生命科学名著

细菌分子遗传学（原书第五版）

Molecular Genetics of Bacteria (Fifth Edition)

〔英〕J.W. Dale S.F. Park 著

王凤阳 等 译



科学出版社

(Q-3152.31)

生命科学名著·典藏版



呈现国际学术精华，引进国际经典名著

细胞生物学精要（原书第三版）	植物生物学
分子生物学（原书第五版）	表观遗传学
基因的分子生物学（原书第七版）	蛋白质物理
Brock微生物生物学（上下册）	进化
遗传学：基因和基因组分析（第八版）	细胞
细菌分子遗传学（原书第五版）	藻类学
神经生物学：从神经元到脑（原书第5版）	基因组3
人类分子遗传学	森林遗传学
衰老分子生物学	基因组学概论
发育的原理（影印版）	癌生物学
糖生物学概述（原书第三版）	Lewin 基因X
结构生物学：从原子到生命	糖生物学基础
Brock微生物生物学（影印）（原书第十版）	微生物基因组
生物化学--基础理论与临床（原书第六版）	兽医微生物学
细胞生理学手册——膜生物物理学精要（下册）（原书第四版）	
细胞生理学手册——膜生物物理学精要（上册）（原书第四版）	



科学出版中心 生物分社
联系电话：010-64012501
E-mail：lifescience@mail.sciencep.com
网址：http://www.lifescience.com.cn

销售分类建议：分子生物学；遗传学；微生物学



赛拉艾芙
生命科学订阅号

www.sciencep.com



ISBN 978-7-03-047485-8

定价（全套）：4500.00 元

生命科学名著·典藏版

细菌分子遗传学

（原书第五版）

Molecular Genetics of Bacteria

(Fifth Edition)

〔英〕 J. W. Dale　S. F. Park　著

王凤阳 等　译

科 学 出 版 社

北　京

图字：01-2011-2240 号

内 容 简 介

本系列丛书均选择生命科学领域经久不衰的经典名著，作者均为国际一流专家，堪称各个专业领域的国际第一书。每一本书的更新都紧跟学科发展，更加适合当前的学习和研究。

本丛书包括《癌生物学》、《分子生物学》、《神经生物学：从神经元到脑》、《表观遗传学》、《基因的分子生物学（第七版）》、《细胞生物学精要（原书第三版）》、《结构生物学：从原子到生命》等30本经典著作。

图书在版编目（CIP）数据

生命科学名著：典藏版/（美）温伯格（Weinberg, R. A.）等编著；詹启敏等译. —北京：科学出版社，2016
ISBN 978-7-03-047485-8

Ⅰ.①生… Ⅱ.①温…②詹… Ⅲ. ①生命科学–研究 Ⅳ.①Q1-0

中国版本图书馆 CIP 数据核字(2016)第 043879 号

责任编辑：王 静 李 悦
责任印制：张 伟 / 封面设计：刘新新

科学出版社 出版
北京东黄城根北街 16 号
邮政编码：100717
http://www.sciencep.com

北京厚诚则铭印刷科技有限公司 印刷
科学出版社发行 各地新华书店经销
*
2016 年 7 月第 一 版 开本：787×1092 1/16
2016 年 7 月第一次印刷 印张：1431 3/4
字数：33 950 000

定价：4500.00 元

(如有印装质量问题，我社负责调换)

《细菌分子遗传学》译者名单

主　译　王凤阳　杜　丽　何洪彬

副主译　崔　克　张英霞　胡日查　廖承红　王洪梅
杨宏军　侯佩莉　孙玉成　满初日噶
杨雨辉　范泉水　成　鹰　张珈宁　雷　明
张冬琳　张晓茹　匡文华　焦寒伟　郝永昌
荣　辉　贾晓晓　史巧芸　郭莳雨　祁　超
郭建全　邱　薇　张富强　黄雪英　胡金燕

审　校　范泉水　杨雨辉

序

众所周知，细菌分子遗传学是微生物学的一个重要分支。其内容不仅涉及细菌的遗传规律，也涵盖了细菌的代谢、生长及致病性等。其研究成果广泛应用于农业、工业和医学领域。近年来，我国的微生物学研究在国家有关部门和全体微生物学工作者的共同努力下，取得了一系列举世瞩目的成绩。但就整体而言，与国际先进水平之间还存在一定的距离，尤其是在细菌分子遗传学方面，还有待加倍努力。

“他山之石可以攻玉”，由 Jeremy W. Dale 等主编的 *Molecular Genetics of Bacteria* 原版书在微生物学领域具有很高的权威性。在第五版中，作者增加了近年来细菌分子遗传学研究的最新成果。该书对从事相关领域教学、科研工作的科技人员来说，无疑是一本十分难得的参考专著。王凤阳、杜丽、何洪彬等年轻学者是兽医微生物学研究领域的后起之秀，以其敏锐的专业洞察力与高度的职业责任感，在完成现任工作的同时，组织以博士为骨干的翻译团队，夜以继日地完成了本书的翻译，这既是他们为繁荣我国微生物学研究所做的贡献，又从一个侧面反映了他们的工作成绩与进步。作为年长于他们的兽医学同行，我由衷地为该书的问世而高兴，同时也为我国兽医事业后继有人而倍感欣慰。在此，我谨向他们表示由衷的祝贺！同时，我也想借此机会诚挚地期望今后有更多、更好的兽医微生物专业相关的编著和译著面世，以尽快迎来我国兽医学繁荣昌盛的春天，为我国的现代化建设做出更多更大的贡献。

军事医学科学院军事兽医研究所研究员　夏咸柱
中　国　工　程　院　院　士

主译自序

从1900年孟德尔遗传定律的重新发现至今，遗传学从经典遗传学、生化遗传学、微生物遗传学、分子遗传学发展到今天的基因组遗传学，研究对象也从果蝇、玉米等发展到真菌、细菌、病毒，以及小鼠等。作为以细菌为研究对象的分子遗传学学科，细菌分子遗传学不仅阐释了遗传学的一些基础原理，而且极大地推动了分子遗传学的发展。

作为从事微生物学研究的科研人员，接触、使用大肠杆菌、噬菌体已经有十几年的时间了。一直想对这些为研究工作做出很大贡献的“可爱的小家伙们”的“品性”做一个全面、系统的解释，可惜没有机会。*Molecular Genetics of Bacteria*（第五版）的翻译工作，可以使我们翻译团队在传统细菌遗传学的基本框架下，加上分子遗传学、基因组学中新的研究成果，更深入、更全面地了解和认识它们。非常希望该书的中文译本出版后，能够为相关领域的读者的理论提高和实际工作提供更多的帮助。

Molecular Genetics of Bacteria（第五版）在以前版本的基础上，对于σ因子、反σ因子、反-反σ因子、与抗生素耐药有关的整合子等概念进行了更深入的阐述，代表了细菌分子遗传学领域在当今世界最前沿的研究水平。本书既可作为高等院校相关专业的研究生和高年级本科生的教材或参考书，也可供生物医学等相关专业的科研人员在工作中参考。

在整个翻译团队的共同努力下，本书历时两载，终于付梓，科学出版社的岳漫宇编辑、李悦编辑、莫结胜编辑功不可没，在此深表谢意。由于水平所限，书中难免有不妥之处，希望读者给予批评指正，以期在重印时更正。

主　译

2013年4月于海口

前　言

在第四版(2004 年出版)的前言中,我们提到,细菌遗传学的革命发端于基因克隆和测序,PCR 和微阵列,以及范围迅速扩大的基因组测序。这无疑会产生一种困惑,在没有忘掉经典遗传学理论的前提下,我们如何才能适应这些新技术及其所提供的激动人心的新信息呢?而这种困惑已经变得越来越尖锐。

的确,许多不再被使用的老的研究方法已经被放到史料中。但这无疑会存在"泼掉洗澡水的同时将孩子扔掉"的危险。我们不仅有必要对于这一学科如何发展到今天保持敏锐的洞察力,而且有必要充分认识这些方法对于在自然环境下细菌遗传学发挥作用时所具有的重要价值。遗传学不仅研究如何认识细菌,也要研究认识细菌过去如何进化,未来如何进化,以及如何适应时刻改变的环境。孤立地看,分子遗传学本质上是简单论,基因组测序和基因表达的全面分析,也仅是提供了基因的目录而已。最终,这些条目还要与生物体的整体行为发生关联,并能够阐释生物体之间如何作用,以及与环境如何作用。

基于此,我们采取了折中的方法,即进一步减少经典遗传学的篇幅,对一些新的技术进展和这些方法对于细菌行为研究的重要意义着以更多的笔墨。

一个需要进一步予以解释的概念是,何为细菌?显然,有两种不同类型的原核生物:细菌(bacteria)和古菌(archea)。本书的大部分内容涉及固有的分类学观点的细菌,但有些方面也与古菌有关,尤其是当我们考虑到原核生物和真核生物的不同时,偶尔使有些内容难以区分,对此,我们表示歉意。

和较早的版本一样,本书内容的取舍完全是作者的个人观点。非常希望本书能够适于非专业读者阅读,并且能够成功地引导他们进入到激动人心和迅速发展的分子遗传学领域以及令人陶醉的细菌世界。

Jeremy W. Dale
Simon F. Park

目　录

1 核酸的结构和功能

在阅读本书时，基本的分子生物学知识，尤其是对于核酸和蛋白质的结构和合成的了解是必要的。因此，本章仅对一些相关的概念进行了回顾，同时强调了那些对理解后面章节内容尤为必要的某些重要特性。

1.1 核酸结构

1.1.1 DNA

细菌的遗传物质为双链 DNA，尽管噬菌体（感染细菌的病毒，见第 4 章）的遗传物质可能是双链 DNA、单链 DNA 或 RNA。DNA 的组成包括 2′-脱氧核糖（经磷酸基团连接形成骨架）和 4 种杂环的碱基：两种嘌呤[腺嘌呤（A）、鸟嘌呤（G）]和两种嘧啶[胸腺嘧啶（T）、胞嘧啶（C）]（图 1.1）。在一个脱氧核糖的 5′位置与下一个脱氧核糖的 3′位置之间，糖基经磷酸二酯键连接（图 1.2），而某一个碱基连接到脱氧核糖的 1′位置。恰恰是这 4 种碱基的排序携带了遗传信息。

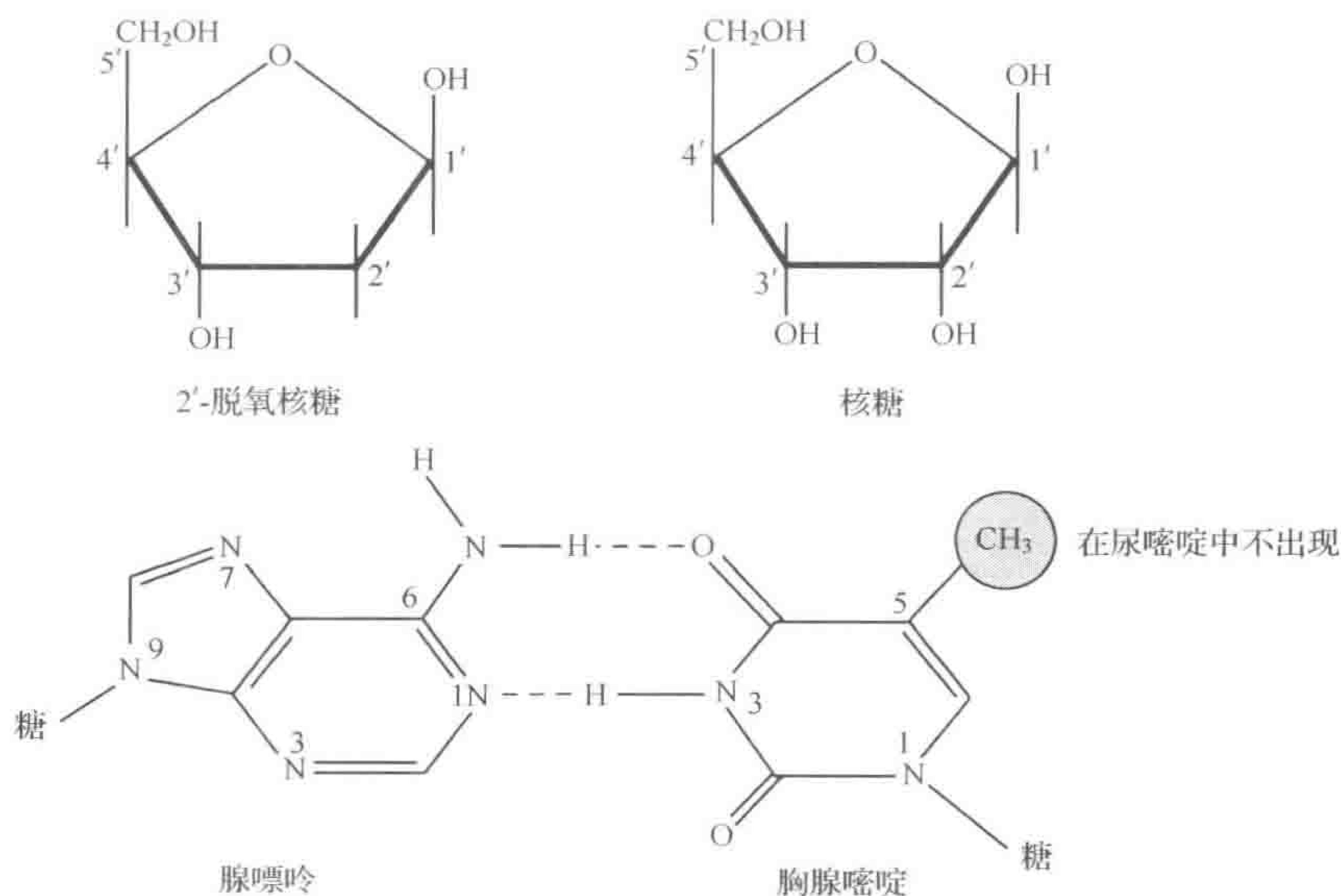


图 1.1　DNA 和 RNA 基本组成结构

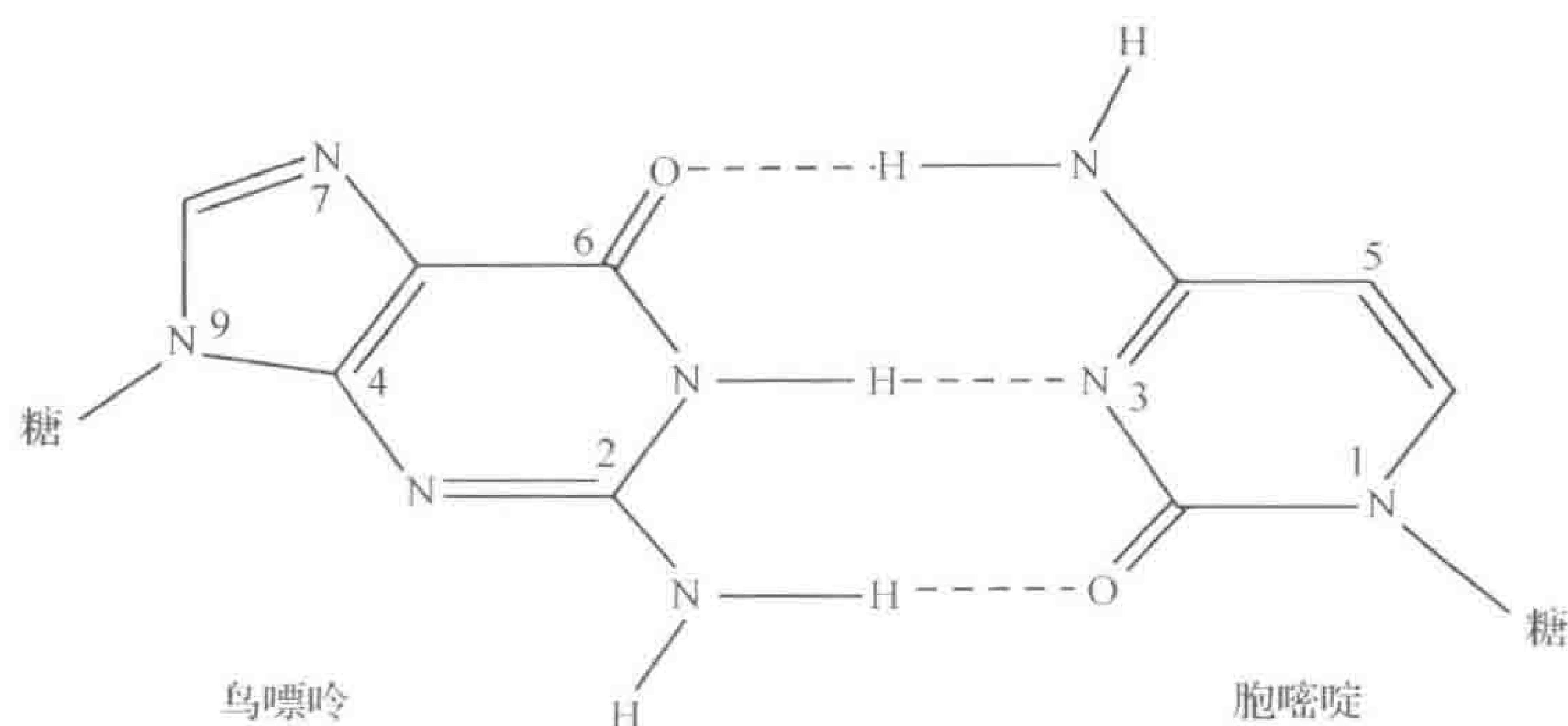


图 1.1 DNA 和 RNA 基本组成结构(续图)
RNA 含有核糖而非脱氧核糖,尿嘧啶而非胸腺嘧啶

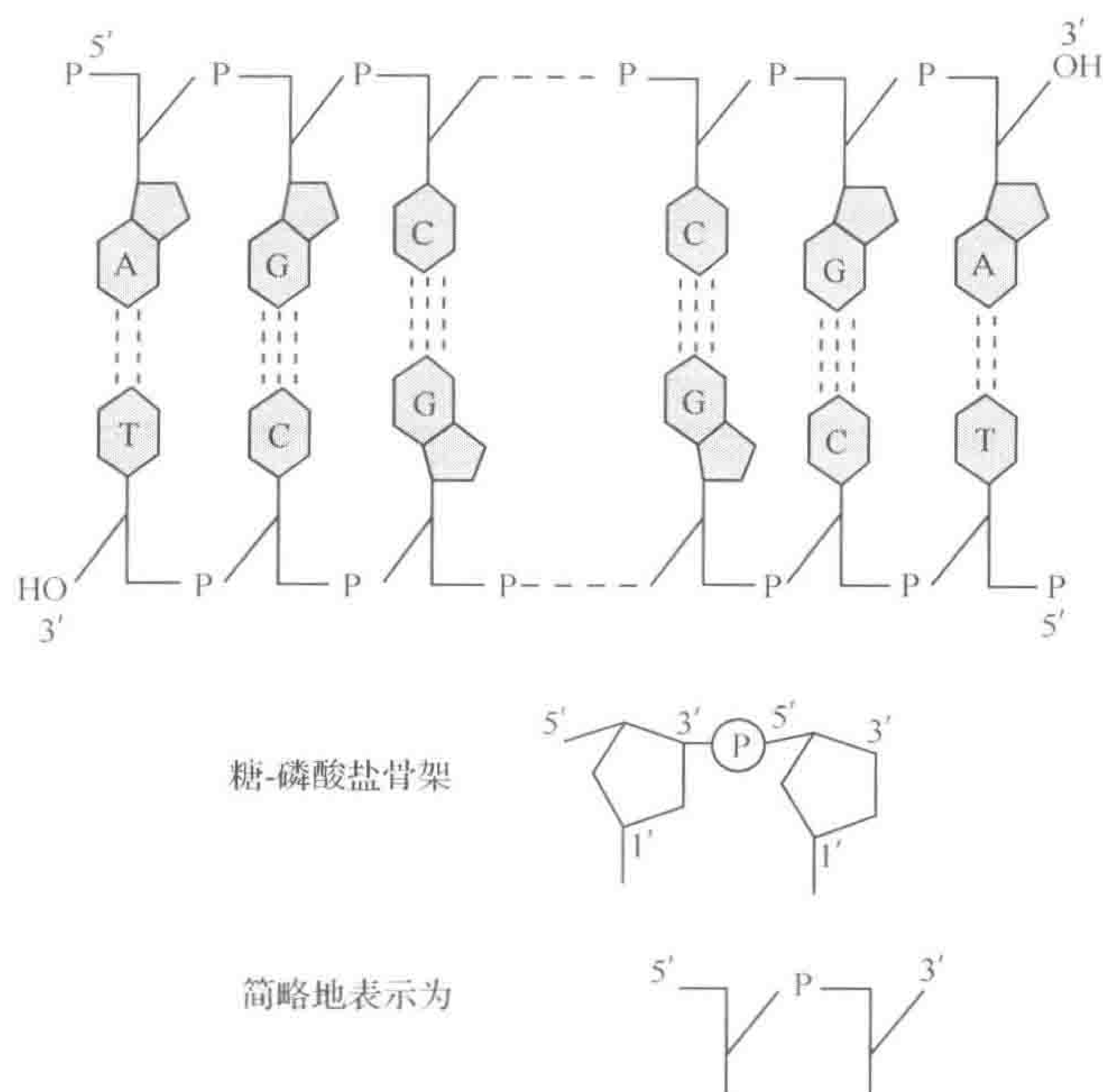


图 1.2 DNA 结构图解

众所周知,两股链以双螺旋形式相互扭转,碱基位于中心,糖-磷酸盐骨架处于外侧。碱基间的氢键将两股链连接在一起。当腺嘌呤和胸腺嘧啶配对、鸟嘌呤和胞嘧啶配对时,碱基排布使双螺旋始终保持正确构象。一条链由另一条链的镜像组成,这两条链被称为互补。值得注意的是,嘌呤分子大小大于嘧啶分子,这种排布涉及每个位置上的一种嘌呤对应一种嘧啶,因此链间的距离保持恒定。

1.1.2 RNA

与 DNA 的结构不同,RNA 包含核糖,却不包含脱氧核糖,包含尿嘧啶,却不包含胸腺嘧啶(图 1.1)。通常情况下,仅由于没有合成互补链,RNA 被描述为单链。RNA 没有

阻止其形成双螺旋的固有结构:一条 RNA 链可以与一条互补的 RNA 链配对(杂交),或是与一条互补的 DNA 链配对。一条单链 RNA 甚至可以自身折叠形成双链区,特别是转移 RNA(tRNA)和核糖体 RNA(rRNA)都能形成碱基配对区的复杂方式。RNA 通过碱基配对形成的二级和三级结构,也能影响基因的表达,这将在第 3 章进行更详细的讨论。

1.1.3 疏水作用

尽管遗传学家强调两条 DNA 链间氢键的重要性,但这不是影响 DNA 结构的唯一作用力。碱基的疏水性使之趋向于疏离水相环境。通过碱基的堆叠可以部分达到这一效果(图 1.3)。两条链中碱基间疏水作用使双链结构更加稳定。氢键不仅将两条链连在一起,而且使互补碱基紧密相连,发挥其疏水作用。然而,碱基间的氢键尤为重要,更主要的原因是其确保了两条链间碱基配对的特异性。

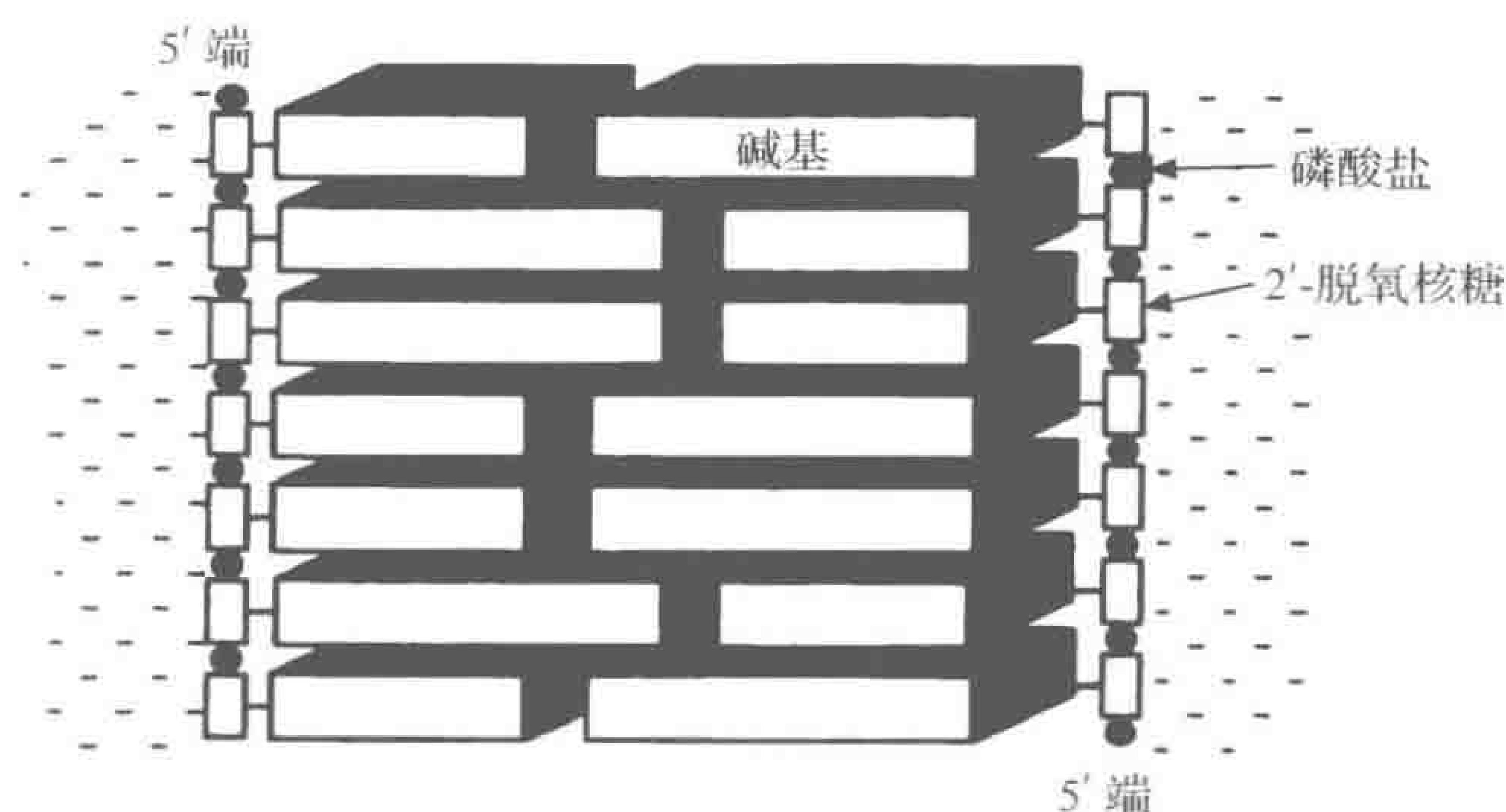


图 1.3 DNA 中碱基疏水作用
疏水碱基堆叠在双螺旋的中心,降低它们与水的联系

虽然碱基是疏水的,在水中很难溶解,但核苷酸易溶于水,主要因其骨架的亲水性,特别是带负电荷的磷酸基高度密集。疏水的碱基位于中心,与水隔离,亲水的磷酸基暴露在外,这将有利于双螺旋结构的形成。

1.1.4 双螺旋的不同形式

对 DNA 结构的全面认识是相当复杂的,应考虑周围水的自身作用,其他溶质和溶剂的影响作用。因此,从某种程度上来讲,DNA 结构可以根据这些条件发生改变。在体外条件下,发现了两种主要的结构形式。Watson 和 Crick 结构指的是 B 构象,即每圈螺旋含 10 个碱基对的右手螺旋(图 1.4)。在特定条件下,分离到的 DNA 可以采取替代形式,被称为 A 构象,即更加紧密的、每圈螺旋含 11 个碱基对的右手螺旋。在细胞内,DNA 更倾向于形成 B 构象,但每圈螺旋大约含 10.4 个碱基对(欠旋)。

特定的 DNA 序列,尤其是那些包含交替的 G 和 C 残基的序列更趋向于形成左手螺旋,即 Z 构象(由于糖-磷酸盐骨架是一个 Z 型结构,而不是 B 构象通常显示的曲线型)。虽然最初发现合成的寡核苷酸为 Z 构象,细胞中发现的天然 DNA 可在短距离内或暂时

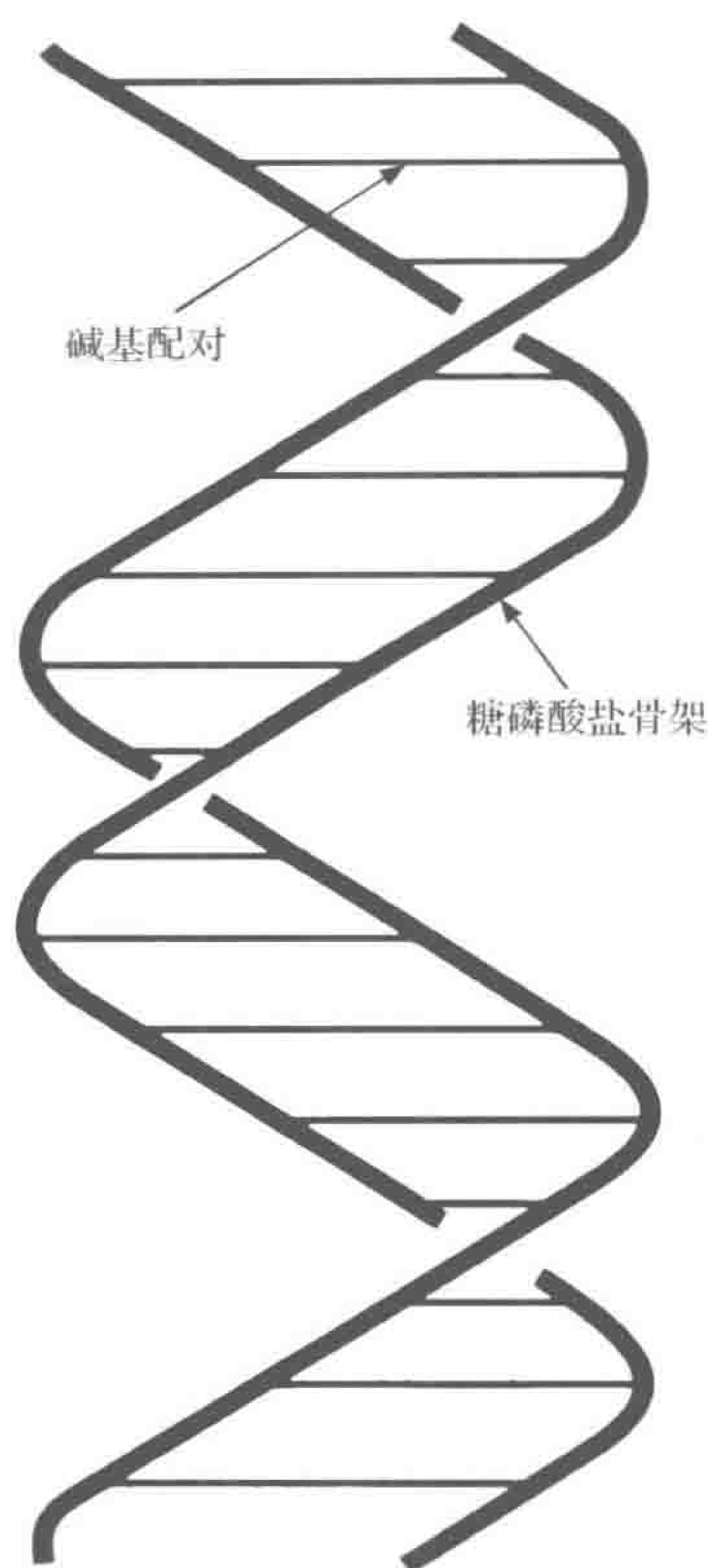


图 1.4　B 构象 DNA 结构图解

这两条反向平行的糖磷酸链构成右手双螺旋，伴随中心的碱基对，通过疏水作用和氢键连接相结合

采用左手螺旋结构。左手螺旋结构向右手螺旋结构的转换对于该区域内基因的表达具有重要影响。

1.1.5　超螺旋

在细胞内，DNA 双螺旋进一步缠绕形成超螺旋。图 1.5 是一个超螺旋的简单演示，学生可以很容易地进行操作。将一条纸带的末端完整旋转一圈（纸带的同一面的两个末端均面向自身），现在看起来就像图 1.5(a)显示的形式，然后将两末端放在一起，构象将呈现出图 1.5(b)中的样子，即超螺旋的一种简单形式。值得注意的是，不仅是一条纸带变成超螺旋，而且卷曲程度也发生了改变（在例子中没有出现完全卷曲）。如果将两端拉伸，双螺旋的卷曲也不会完全消失，只是会变为一种不同的形式。如果再次将两端拉开，将会变回到图 1.5(a)所显示的形式。

与螺旋相关的 3 个参数包括：缠绕数(T)、链环数(L)、超螺旋数(W)。缠绕数是双螺旋的螺旋数，而超螺旋数（即超螺旋程度的估量）可看作双螺旋在一特定方向自身交叉的次数。由于构象的变化，这两个参数也发生变化。在图 1.5(a)中，有一次缠绕($T=1$)但无超螺旋($W=0$)，而在图 1.5(b)中，没有缠绕($T=0$)，双螺旋自身交叉一次($W=1$)。作

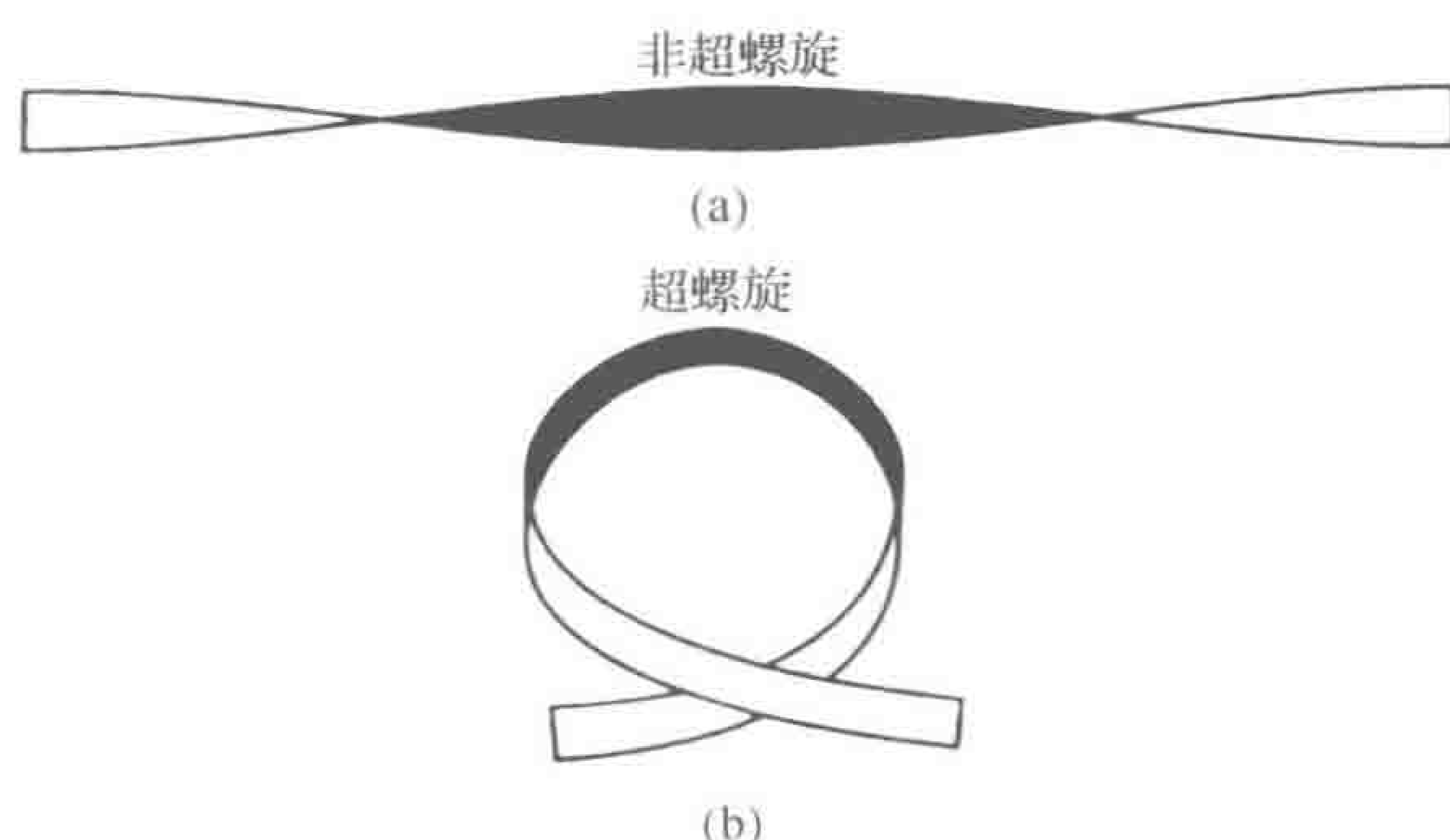


图 1.5 扭转和超螺旋间的联系

(a) 一个完全扭转的条带，没有超螺旋；(b) 相同的条带，形成一个超螺旋，现在条带没有扭转

为双螺旋完全缠绕度量单位的链环数等于这两个参数之和，即 $L=T+W$。

如果条带的两端不能自由旋转，链环数将保持不变。所考虑的大部分 DNA 分子是环形的，因此不包括可旋转的末端。除非 DNA 出现断裂，缠绕中的任何改变都将被超螺旋的改变所平衡，反之亦然(图 1.6)。

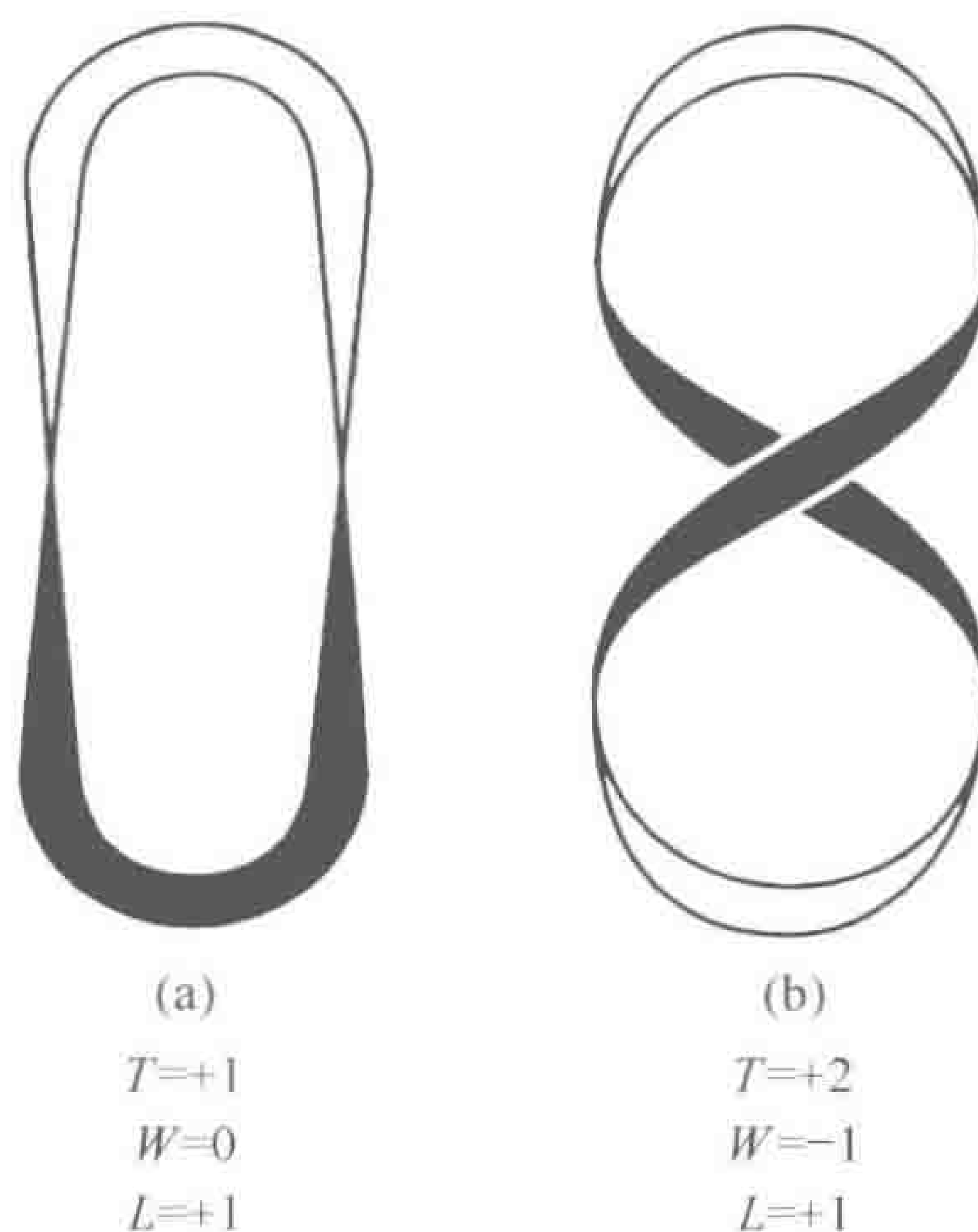


图 1.6 一个圆环分子的超螺旋

在(a)中，这个“分子”有一次扭转没有超螺旋，该分子一端旋转成(b)形成负超螺旋且增加了扭转数，因此链环数(L)保持不变，这两种形式彼此转换没有破坏环状结构

图 1.6(a)中的条带(或 DNA 分子)不是超螺旋($W=0$)，但包含一个完整的缠绕($T=+1$)，链环数(L)是$+1$。在图 1.6(b)中，通过旋转结构的一端，整个条带的形态发生改变(如发生一定程度的超螺旋)。该条带自身交错一次，按照规定，定义在该方向的交叉为负值，因此 $W=-1$。同时，缠绕数发生改变，现在发生了两个完全的缠绕，所以 $T=+2$。因为 $L=T+W$，所以发现 L 保持不变($+1$)。

通过转动一端，图 1.6 中显示的两个结构在圆环没有断开的状态下实现相互转换。对于一个完整的圆环条带，可以同时改变缠绕数和扭转数。在超螺旋中任何改变都将引起缠绕数的补偿性改变(反之亦然)，使得链环数保持恒定。环状 DNA 的链环数只有在

DNA 链断裂和重新连接时可能发生改变，如通过拓扑异构酶的作用(见下文)。

细菌的 DNA 通常是负超螺旋。另一种表述即是 DNA 是欠旋的，如果 DNA 没有超螺旋，双螺旋扭曲程度会比松弛的线性 DNA 低。如果出现裂口(也就是一股断裂，使之自由转动)，DNA 松弛成一个开环，即非超螺旋的形式。在细胞裂解过程中，染色体 DNA 常被割裂成线性片段，但在通常情况下，由于细菌质粒足够小，可以分离到没有被破坏的超螺旋形式(见第 5 章)。

DNA 紧密的超螺旋结构对染色体的形成具有重要意义，因为在伸展状态下的 DNA 会比菌体细胞自身(约 1mm)长 1000 倍。换句话说，一个细菌含 4 种基因的操纵子，如果以非超螺旋的 B 构象形式存在，可以从细胞的一端伸展到另一端。超螺旋仅是个开始，因为细菌染色体包含大量的超螺旋环，它们排布于中心，产生一个高度密集和有序的结构，即类核。DNA 超螺旋(和另一些结构特征)在基因表达调控方面也很重要(见第 3 章)。

拓扑异构酶的功能

细菌 DNA 的超螺旋并非如图 1.6 中所显示的那样通过物理扭转环状分子的方式形成，而是细菌应用 DNA 拓扑异构酶控制 DNA 链的断裂和再连接，从 DNA 引入(或消除)超螺旋。

DNA 拓扑异构酶分为两类。拓扑异构酶Ⅰ通过断裂一条链，并使另一条链穿过缺口后再连接缺口，作用于 DNA 片段。由于两股链彼此交叉的次数增加，链环数增加了 1，使得 T 或 W 中一个值增加。大肠杆菌(*Escherichia coli*)拓扑异构酶Ⅰ只作用于负超螺旋 DNA，W 值的增大意味着负超螺旋的程度减弱(DNA 变得松弛)。

拓扑异构酶Ⅱ使 DNA 双链断裂，另一螺旋穿过缺口。在图 1.7 中，从一个非超螺旋的环形结构(A 结构)开始，将右手和左手环的中心移动，彼此交叉形成 B 结构。尽管这乍看起来是超螺旋，但实际并非如此。这两个交叉处方向相反，因此相互抵消(数学上，上面的交叉点 $W=-1$，下面的交叉点 $W=+1$，所以总体上 $W=0$)。如果螺旋的两条链在 L 与 M 点间断裂，压在下面的链(X-Y)穿过该缺口，在 L 和 M 点再将链重新连接，就形成了 C 结构。现在，两交叠处方向相同，所以该结构为超螺旋($W=-2$)，链环数相应减小。该酶引入了负超螺旋，同时降低了双螺旋的旋转。这类酶的一个重要例子是 DNA 旋转酶，它可以将负超螺旋引入到新复制的 DNA 中。

1.1.6 变性与杂交

由于 DNA 双链仅由非共价键连接，在实验室条件下，它们很容易分离，如升高温度或提高 pH。DNA 双链的分离，即变性，常可以逆转。降低温度或 pH 可使互补的 DNA 序列间重新形成氢键，即退火(图 1.8)。如果不同来源的 DNA 分子发生变性、混合、退火，相似的 DNA 序列之间很可能形成氢键(杂交)。这是使用 DNA 探针发现特异性 DNA 序列的理论基础。通过改变退火条件(或随后的洗涤)可以调整杂交的特异性。较高的温度或偏低的离子强度会增强杂交的严格性。严格性较高的杂交常用于发现紧密相关的序列，或分辨序列间很小的差别，而低严格条件常用于发现与探针差别更大的序列。

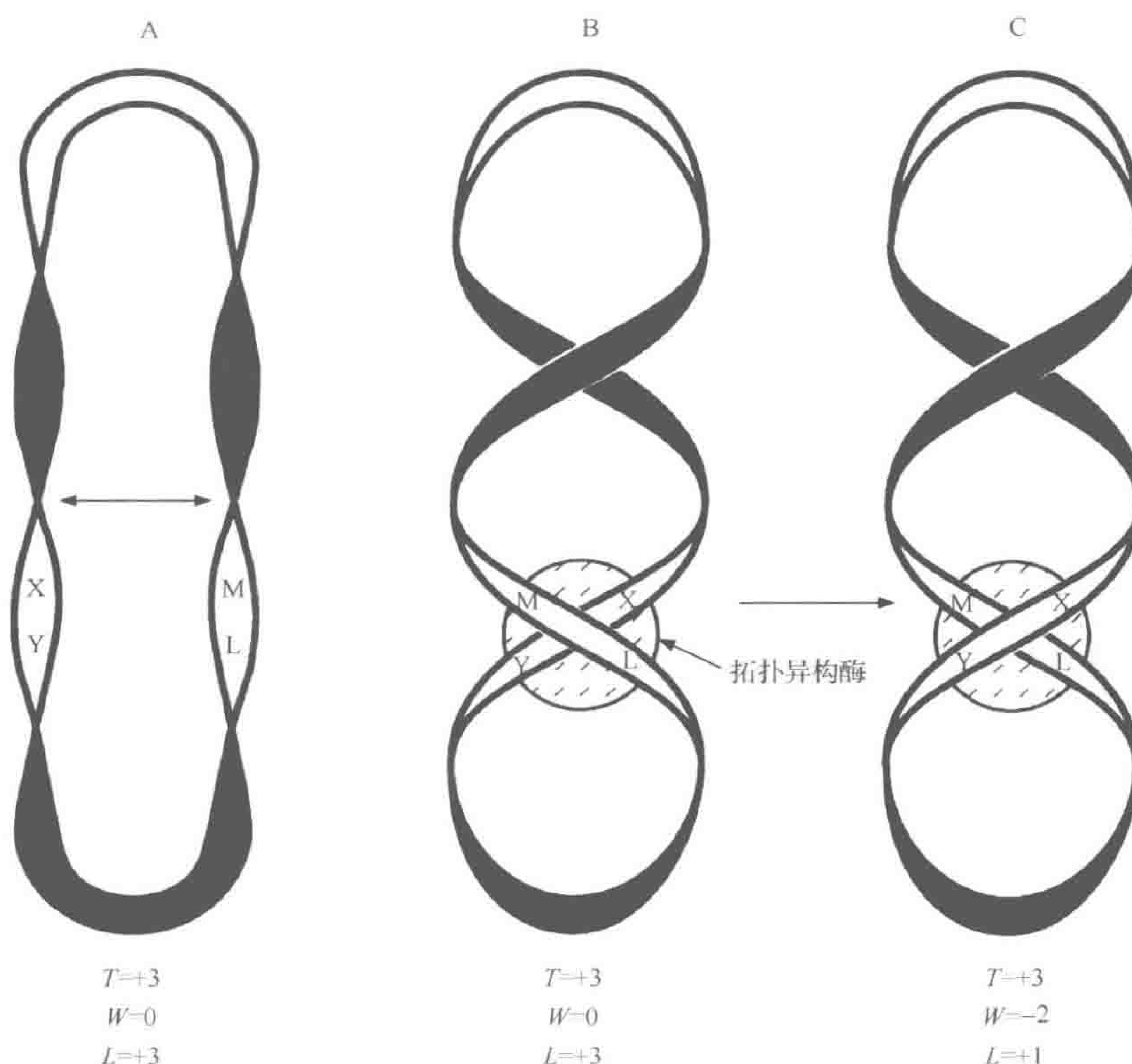


图 1.7 拓扑异构酶Ⅱ功能

A 结构不是超螺旋，按箭头所指两边弯曲转变成 B 结构；B 结构不是超螺旋，两交汇点标记截然相反而相互抵消；拓扑异构酶使双螺旋中 L 与 M 间断裂，穿过 X-Y 区域的间隙再修复 L 与 M 间的区域，这样就改变了在那一段的 W 值，所以结构 C 现在为负超螺旋

这一技术是现代分子生物学的一个重要组成部分，在后面的章节中会涉及其一些应用。

DNA 双链局部的暂时分离也是复制和转录过程的一个必不可少的部分。值得注意的是，有 3 个氢键连接鸟嘌呤和胞嘧啶，而腺嘌呤-胸腺嘧啶配对仅有两个氢键。因此，在高 G+C 含量的区域，DNA 双链的连接更稳定。该区域更难变性，却更容易退火。链间碱基组成对于两条核苷酸链分离难易程度的影响可能在控制 RNA 合成的起始等过程中扮演着重要的角色，如高 A-T 区域可能促进 DNA 链的起始分离（见第 3 章）。

1.1.7 核苷酸链的方向

双螺旋另一个更显著的特征是每条链都有基于糖-磷酸盐骨架连接指向的方向。每个磷酸基团连接一个糖基的 5′端和下一个脱氧核糖的 3′端。在图 1.2 中，上面的链在左手端尾部有一个游离的 5′基团，在右手端尾部有一个 3′-OH。因此，可将其称为 5′→3′方向（从左到右）。反之，下方的这条链从右到左称为 5′→3′方向。

按照惯例，如果表述一条 DNA（或 RNA）单链，则从左到右读作从 5′→3′（除非特别说明）。如果表述的是两条链，则上方的链从左到右读作 5′→3′。

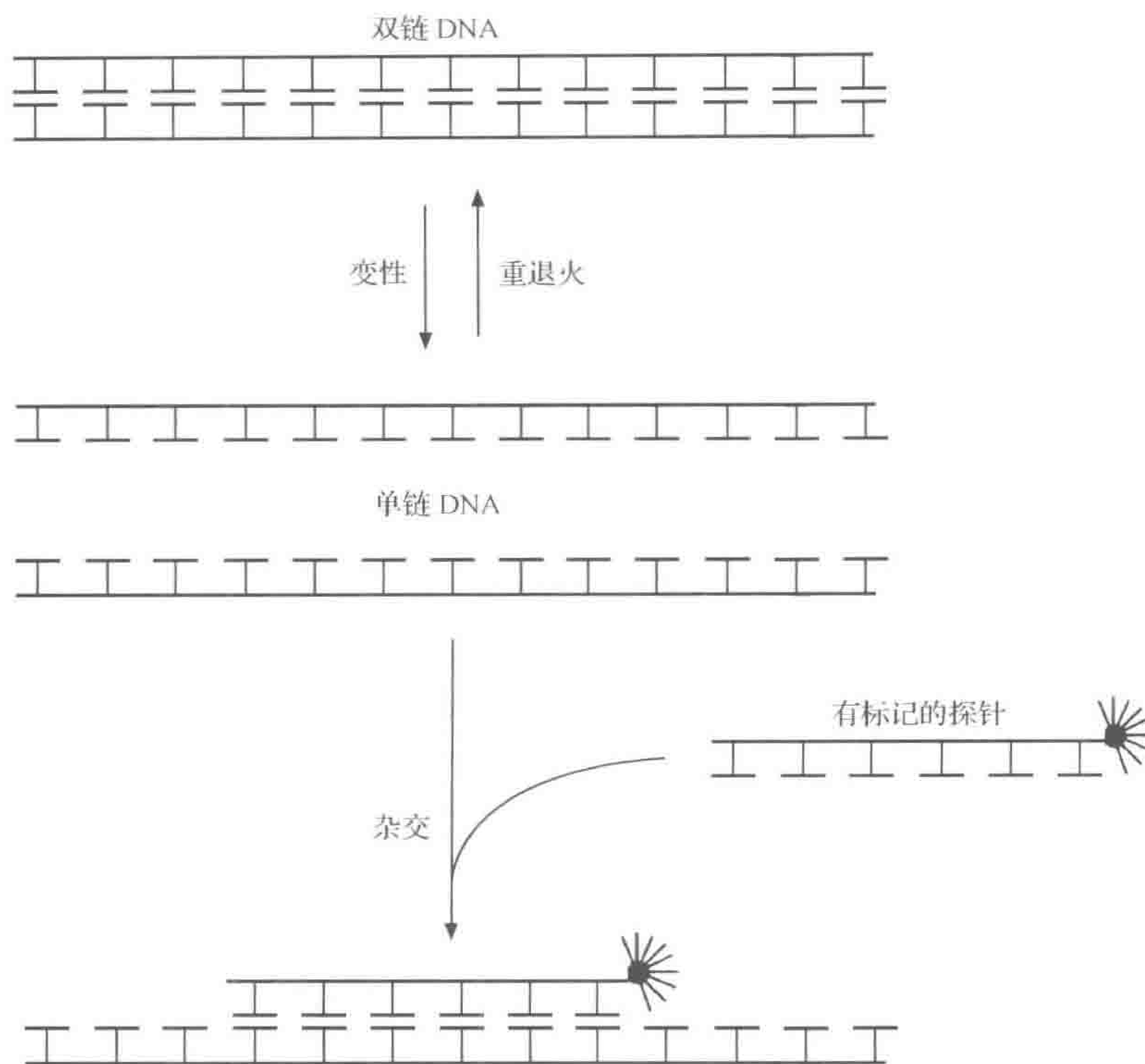


图 1.8 DNA 的变性和杂交

所有的核酸链都是从 5′→3′方向合成，即通过依次将核苷酸添加到前一个核苷酸的 3′-OH，新链得以延伸。5′核苷三磷酸(ATP、GTP、CTP 和 UTP 用于 RNA；dATP、dGTP、dCTP 和 dTTP 用于 DNA)作为底物提供磷酸基团用于连接。

1.2 DNA 复制

一条 DNA 链可作为模板用于合成一条新的核苷酸链，新核苷酸链的每一个碱基与模板链的碱基配对形成氢键(G-C、A-T，RNA 分子为 A-U)。因此，新链与模板链互补。DNA 分子拷贝产生更多的 DNA 的过程称为复制，以 DNA 为模板合成总 RNA 的过程称为转录。

与上面的描述相比，复制更为复杂。图 1.9 概括了其中的一些主要特征。DNA 链的相反极性是一个复杂的因素。新链中的一条(前导链)可在 5′→3′方向连续不间断地合成。对这一合成起主要作用的酶是 DNA 聚合酶Ⅲ。对于另一条新链，总体复制在 3′→5′方向。由于核苷酸仅在 5′→3′方向合成，新的 3′→5′链(后随链)发生“滞后”，即与总体复制方向相反。这一问题可以通过利用 DNA 连接酶将短片段连接在一起形成新链的方式予以解决。

此外，DNA 聚合酶不能启动一个新的 DNA 链，只能延长以前就存在的分子。这一限制不适用于能启动新的核酸合成的 RNA 聚合酶。因此，每个片段由一种特定的 RNA

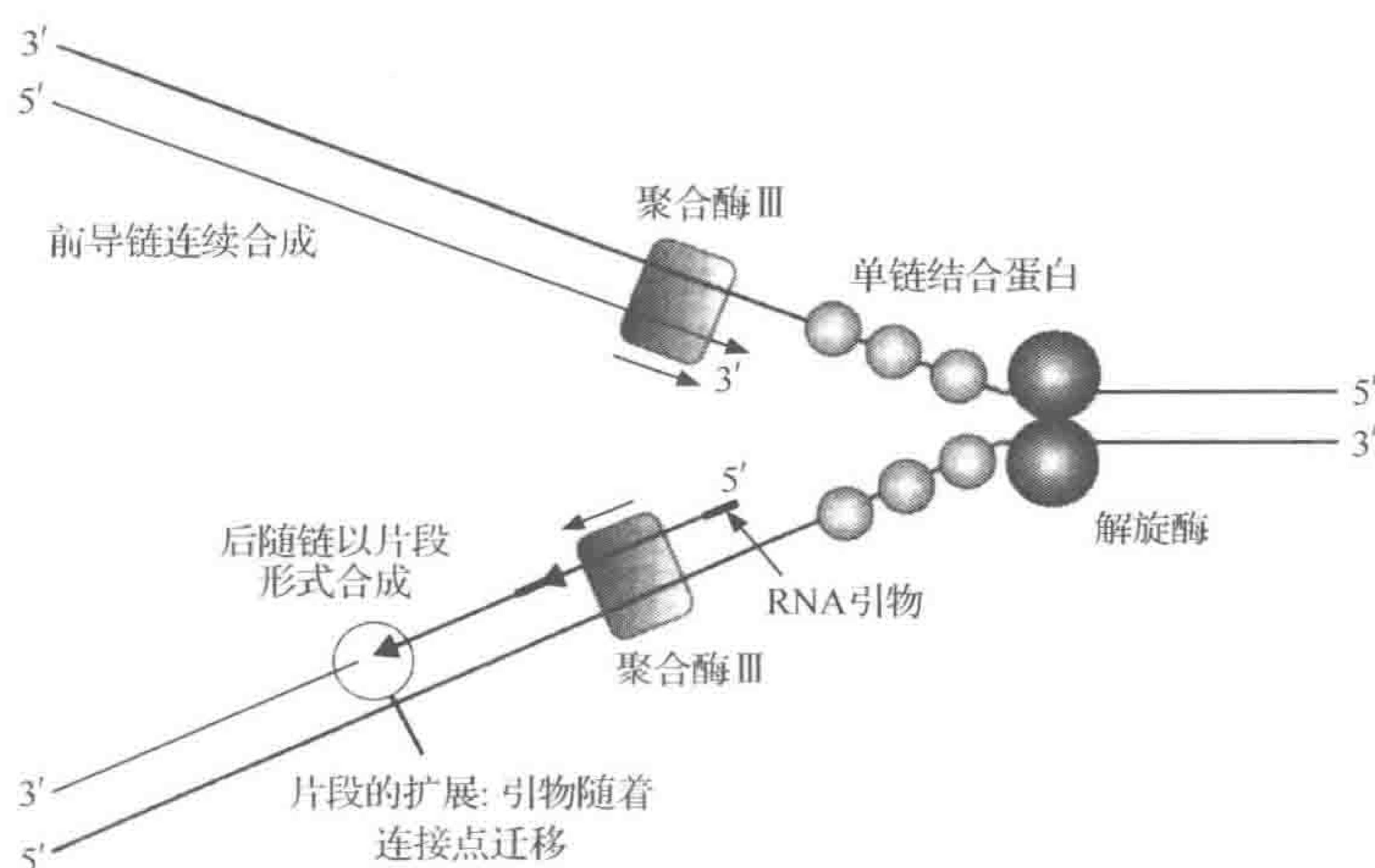


图 1.9 DNA 复制机制简图

注意，图中未显示 DNA 的螺旋结构

聚合酶(引物酶)产生的一小段 RNA 开始，这一 RNA 引物在 DNA 聚合酶Ⅲ的作用下延伸。引物随后被移除，空隙被不同的 DNA 聚合酶(DNA 聚合酶Ⅰ)填满，这种酶具有核酸外切酶以及聚合酶的作用。在缺口填充好后，产生的片段在 DNA 连接酶的作用下连接在一起。

1.2.1 解链和复性

在这些事件发生前，至少在短区域内的两条 DNA 链的分离是必需的。通过解旋酶的作用可以完成 DNA 链的分离，该酶结合在模板链上，并沿着模板链移动。另外一种蛋白(单链 DNA 结合蛋白或 SSB)的结合阻止分离的 DNA 的重新连接。许多 SSB 将结合到 DNA 链上，使 DNA 的一段区域维持延伸的单链 DNA 形式。

两个 DNA 链的彼此缠绕引起另一个难题。细胞内的 DNA 分子不能正常自如地旋转，不仅因为细菌 DNA 通常是环状的。因此，如简化图 1.9 所示，仅通过分离两条链并且合成互补链是不可能产生一对子代分子的。这些链必须解旋分开。如果它们不能自由旋转，在链上某一点的分离将造成其沿着链的过度旋转。除非克服这个问题，否则这些分子很快就会乱成一团(如果不理解这一点，可以拿一些细绳进行实验)。如同前面所说的那样，这个问题的解决需要拓扑异构酶的作用。通过允许双螺旋在复制叉前面展开，可以使链分开复制。在将负超螺旋引入到新复制的 DNA 过程中，拓扑异构酶和 DNA 解旋酶发挥了重要作用。

1.2.2 复制保真；校对

新合成的 DNA 是对模板链的一个精确(互补)复制。这并非简单的源于核苷酸要将其自身排列在正确的位置，而是涉及 DNA 聚合酶可以特异地选择正确排列的核苷酸。

与字面含义相比，除了具有核酸外切酶的作用，大多数 DNA 聚合酶的作用更复杂。

作者发现这样的一个活性：当其向下一个片段延伸时，通过 DNA 聚合酶的 5′→3′核酸外切酶作用可以从冈崎片段上除去 RNA 引物(它可以移除 DNA 链的 5′端碱基)。DNA 聚合酶的第二个核酸外切酶功能增强了复制的保真性：3′→5′核酸外切酶作用切除不断延长的(3′端)DNA 链的核苷酸。这并不像听起来那样不正常，因为 3′→5′核酸外切酶只在 3′端发生不正确配对时才发挥作用。如果 3′端最后的碱基与模板链正确配对，DNA 聚合酶只能通过将核酸增加到 3′端来延长 DNA 链。如果不正确，聚合作用将停止。而 3′→5′核酸外切酶将移除不正确的核苷酸，从而进行接下来的复制(图 1.10)。在第 2 章中将会介绍在向延长的 DNA 链添加核苷酸的过程中发生错误的原因。

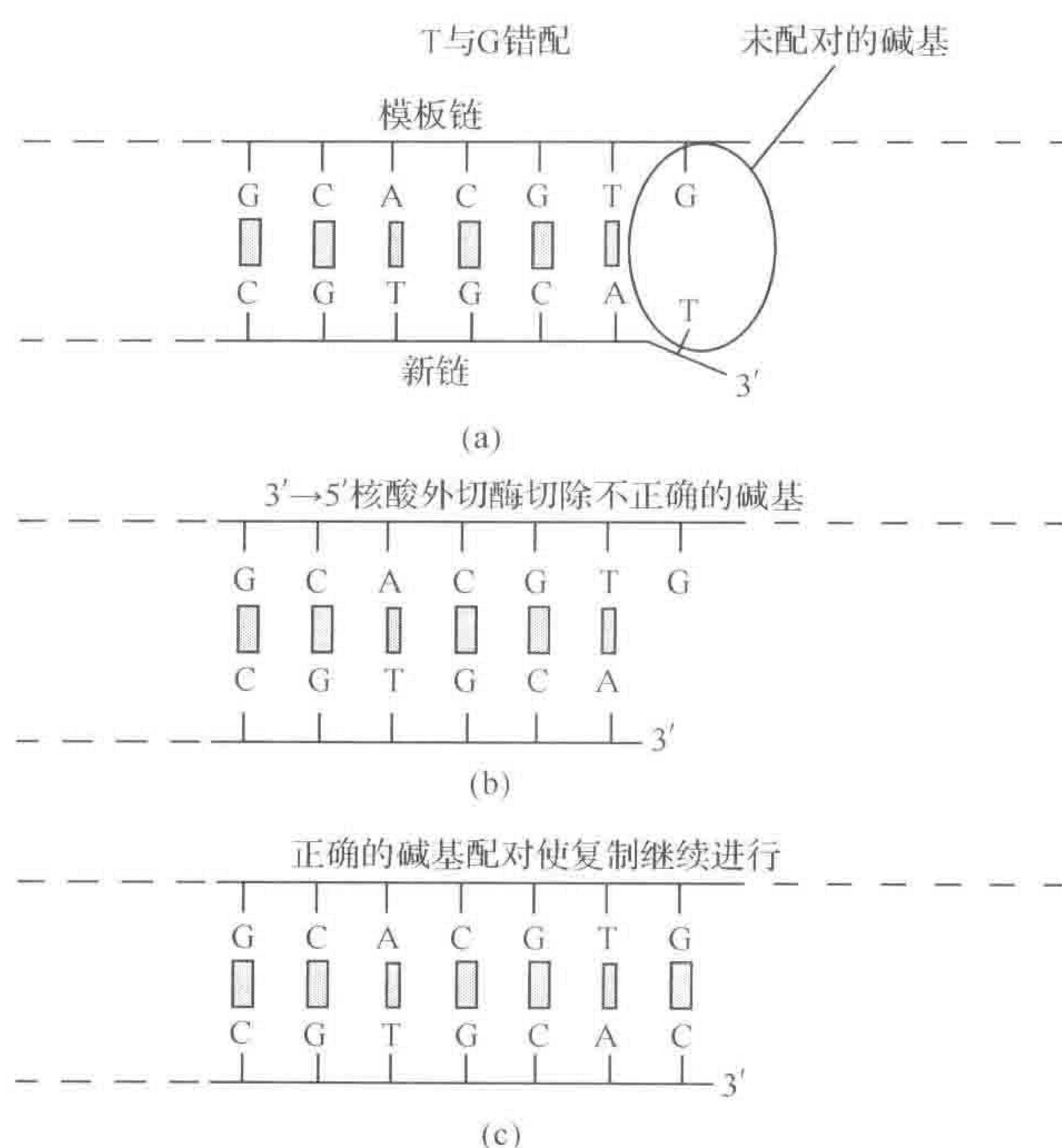


图 1.10　通过校对机制将错误配对的碱基移除

(a) DNA 链延长时加入一个错误的碱基，阻碍 DNA 链的进一步延长；(b)错误配对的碱基被 DNA 聚合酶的 3′→5′核酸外切酶活性移除；(c) 正确的碱基被加到 3′端，DNA 合成继续进行

这种修正错误的机制称为校对或错误校验，使得复制忠实性增加，从而减少了自发突变的比例。然而，这种校对也是有代价的，大量而且广泛的错误校验会减慢复制的速度。DNA 聚合酶本身决定了复制的比例与错误校验程度之间的平衡。一些 DNA 聚合酶不显示有效的校对，会导致更高程度的自发错误的发生。因此，一种细菌发生自发突变的比例是受到进化压力的遗传特点的限制的(至少有一部分)。

DNA 修复机制增强了复制的保真性，这将在本章后面介绍。

1.3　染色体复制和细胞分裂

一般认为，细菌细胞有一个环形的染色体。首先，许多细菌经常包含额外的 DNA 分

子，即质粒。在多数情况下，这些是额外的、可有可无的元件，但对于某些细菌，所有菌株均携带两个或两个以上的不同的质粒，因此两个（或所有的）质粒似乎是正常生长所必需的。其次，并非所有细菌的 DNA 是环形的。有些细菌[特别是链球菌（*Streptomyces*）]有一个线性染色体和（或）线性质粒。这些话题将在第 5 章进一步探讨。

更重要的是，为了确保两个子细胞都获得一份拷贝，细胞在分裂前必须至少有两份完整的染色体拷贝。因此，染色体必须复制以适应细胞分裂周期，这就意味着在细胞周期中的某个时间点，部分染色体已被复制，使得这部分染色体至少含有两份拷贝。

一种细菌染色体的复制通常开始于一个固定的点（复制起点，*oriV*），并且双向进行至与起点相反的终止点（*ter*）（图 1.11）。大肠杆菌的这个过程大约需要 40min。细胞分裂前有至少 20min 的一个时间段，这使复制起始与细胞分裂之间至少有 60min。然而，在有利的条件下，大肠杆菌的生长要比这快得多，可能每隔 20min 就分裂一次。如何使细胞的分裂速度比染色体复制快，而且使得每个子细胞获得一个染色体的完整拷贝？答案在于复制起始的时间。复制起始不是受细胞分裂刺激，而是细胞的一个功能。当这些细胞

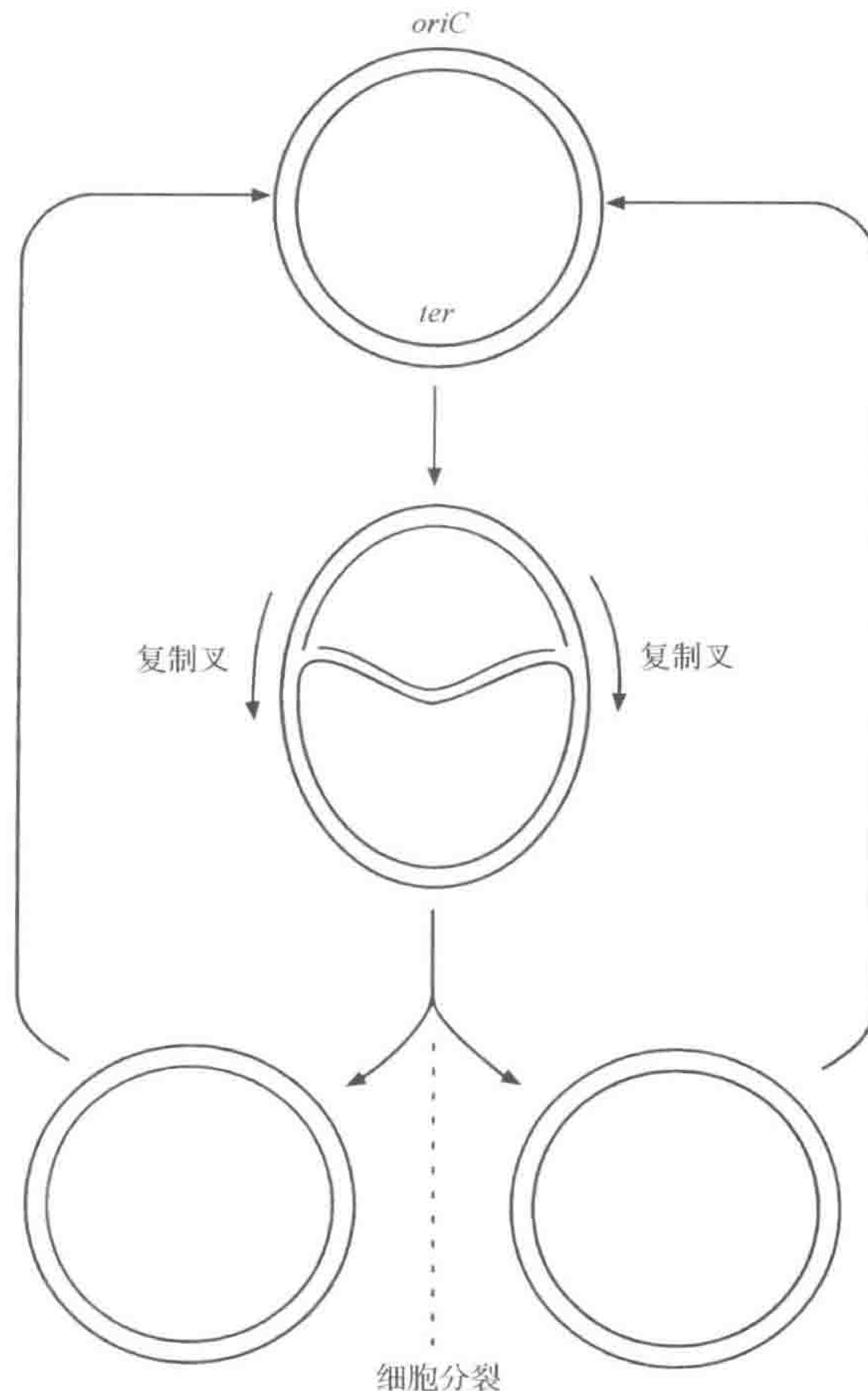


图 1.11 染色体复制

从 *oriC* 位点开始双向复制，到 *ter* 位点结束，生成两个双链分子

快速生长时,有几套复制叉复制染色体,所以,当细胞准备分裂时,不只是有一对完全复制的染色体,这些染色体中的每一个已经被第二对复制叉部分复制(图 1.12)。

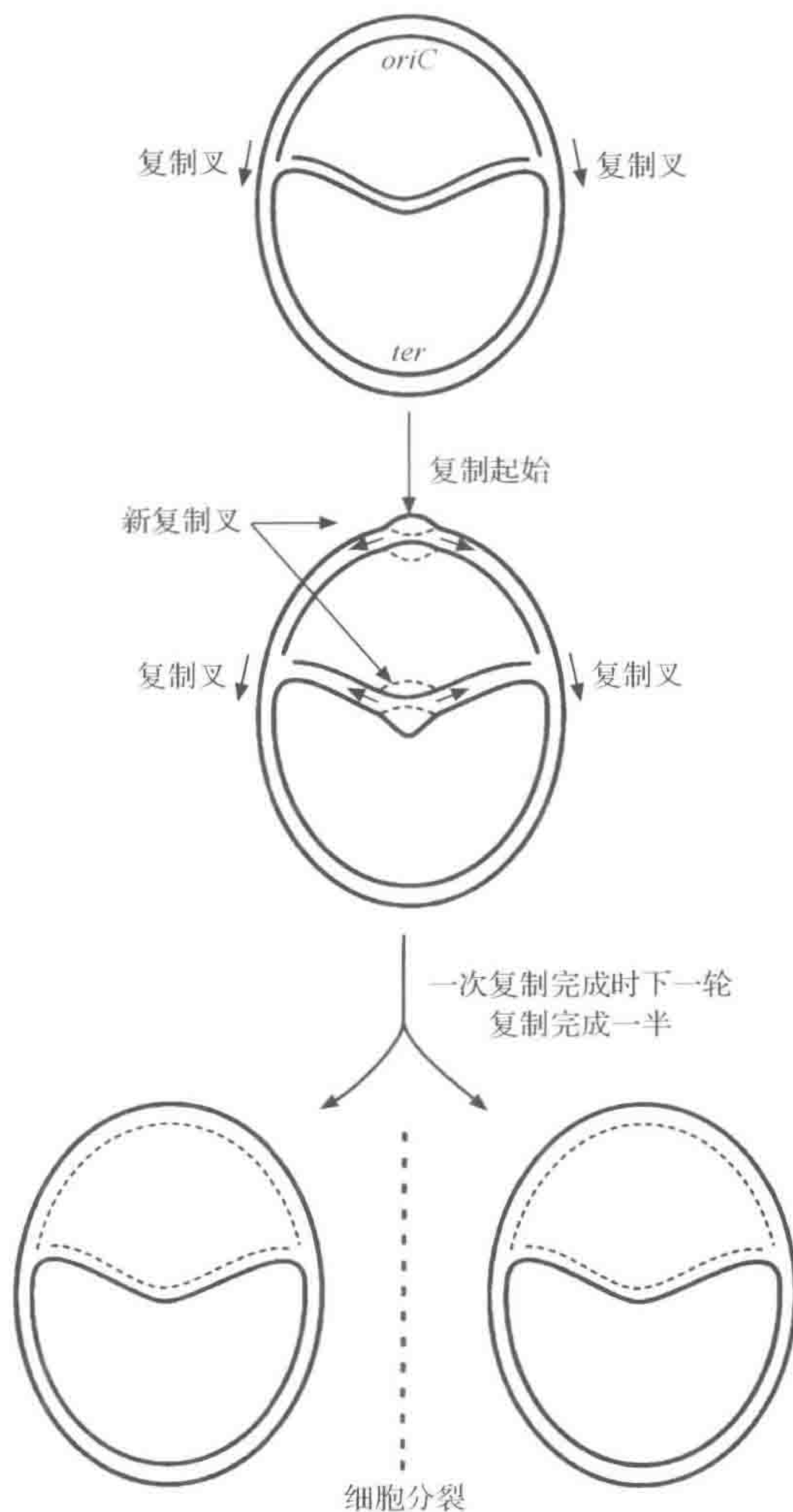


图 1.12 染色体在高生长速率下的复制

当细胞分裂的间隔时间不足以进行染色体复制时,在上一次染色体复制结束前可进行下一轮复制

有两个关键的调节点要考虑:染色体复制完成和随后的细胞分裂之间的联系,以及复制起始的控制。两者都是如此复杂(而且直到目前还没有被完全解释清楚),以至于在这里不能完全得到解释,但能对其中的某些方面进行解释。第一点,一个(简化)模式是染色体复制占据了细胞中点位置的膜的一个区域,并在该点阻止细胞分裂的启动。当复制已经完成时,两个分离的分子可以被分开,移向细胞的两极,从而释放该点使细胞分裂开始。这点将会在第 9 章进行进一步阐述。

第二点,起始的控制更加困难。当细胞达到临界质量时,新一轮复制起始。尝试用这样一种方法,即类似于细胞生长时的一种复制的抑制剂被稀释的方法,来证明临界值的问题,但事实远非如此简单。起始需要一个被称为 DnaA 的蛋白质,DnaA 结合到特定的

DNA 序列(DnaA 盒),复制的起点包含有一定数量的 DnaA 盒。DnaA 蛋白周围缠绕的 DNA 促进链的分离,这对于复制起始是必需的(图 1.13)。有活性的 DnaA 的可用性是复制起始的非常关键的组成部分。然而,控制启动的完整模式更复杂,而且还未完全解释清楚。

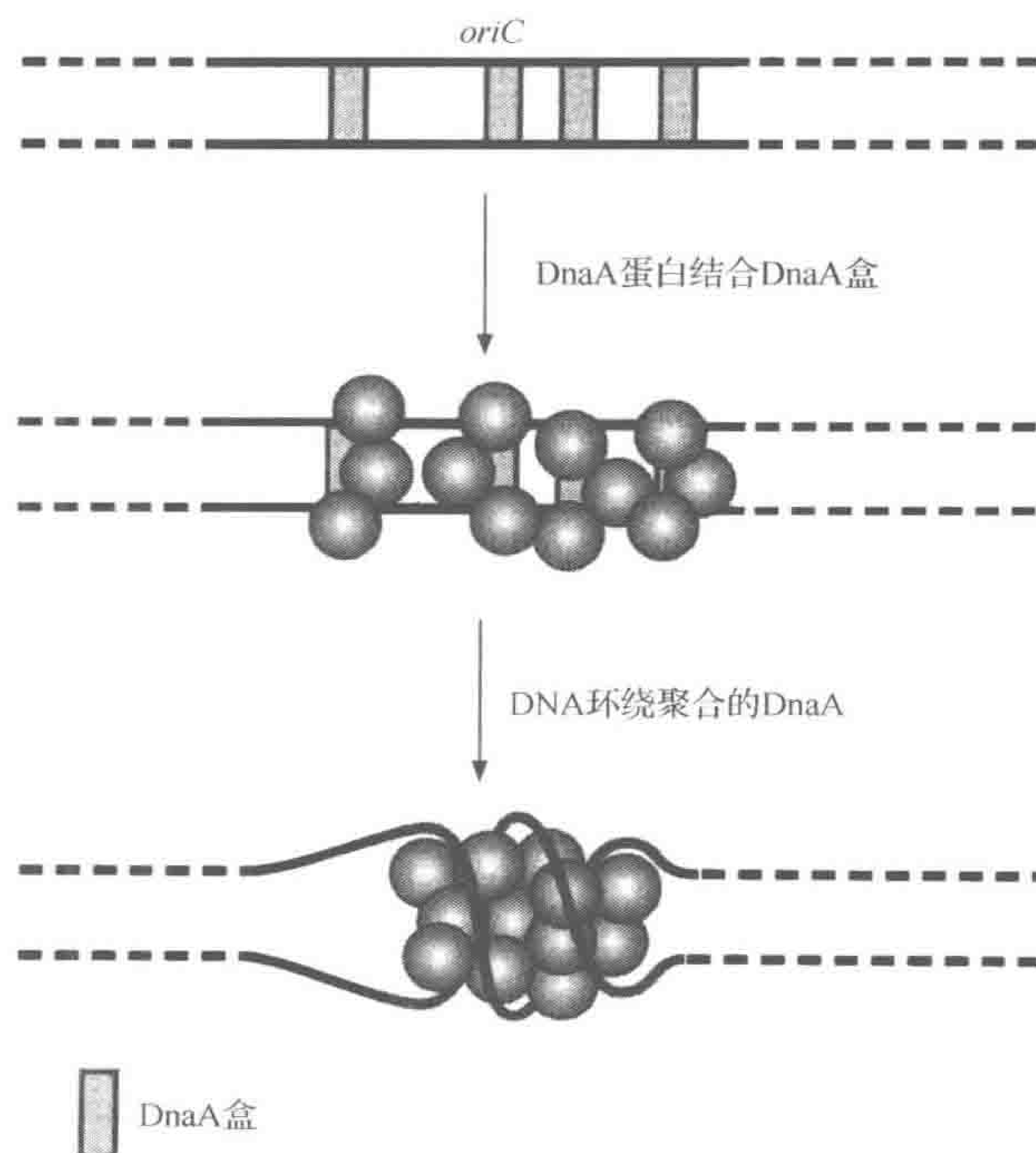


图 1.13 DnaA 蛋白结合在位于 *oriC* 位点的 DnaA 盒上

DnaA 蛋白结合于复制起点处的几个不同位置(DnaA 盒),DNA 缠绕在 DnaA 蛋白的表面以协助打开双链,才能起始复制

1.4 DNA 修复

除了前面所述的这些措施,如增强复制的忠实性,细胞也具有修复受损 DNA 的机制,包括没有校对过程的复制错误以及可能发生在非复制 DNA 上的损伤。

1.4.1 错配修复

最简单的复制错误是错误碱基被纳入了新链。如果发生了这样的情况,又没有启动校对机制,将会导致突变。然而,细胞具有消除这种错配的有效机制,可以用正确的核苷酸取代它们。为了做到这一点,必须知道两条链中哪一条链含有正确的信息。成熟的 DNA 是甲基化的,即它含有额外的甲基,特别是在 GATC 序列中的腺嘌呤存在腺苷的甲基化作用(Dam 甲基化酶)。新链起初不含有这些甲基,所以 DNA 是半甲基化的。至少直到新链甲基化时,新链和旧链可以被区分开。因此,这种机制指的是甲基化定向错配修复,该系统识别错配碱基,去掉一个短的无甲基化区域后填补空缺。

1.4.2 切除修复

其他类型的 DNA 损伤，特别是紫外线辐射形成的嘧啶二聚体（见第 2 章），引起双螺旋结构的变形，能够激活的修复机制称为切除修复。

图 1.14 概括了该机制的本质。这一过程由核酸内切酶启动（由基因 *uvrA*、*uvrB* 和 *uvrC* 编码的一个复杂的酶，这些基因突变引起紫外线抗性降低）。该酶在损伤处的两侧切除 DNA 链，使其 3′-OH 暴露，DNA 聚合酶 Ⅰ 以之为引物填补 DNA 缺口位点之间的短序列（15～20 个碱基）。最后一步是 DNA 连接酶将新修复的链连接到现有的 DNA 上。

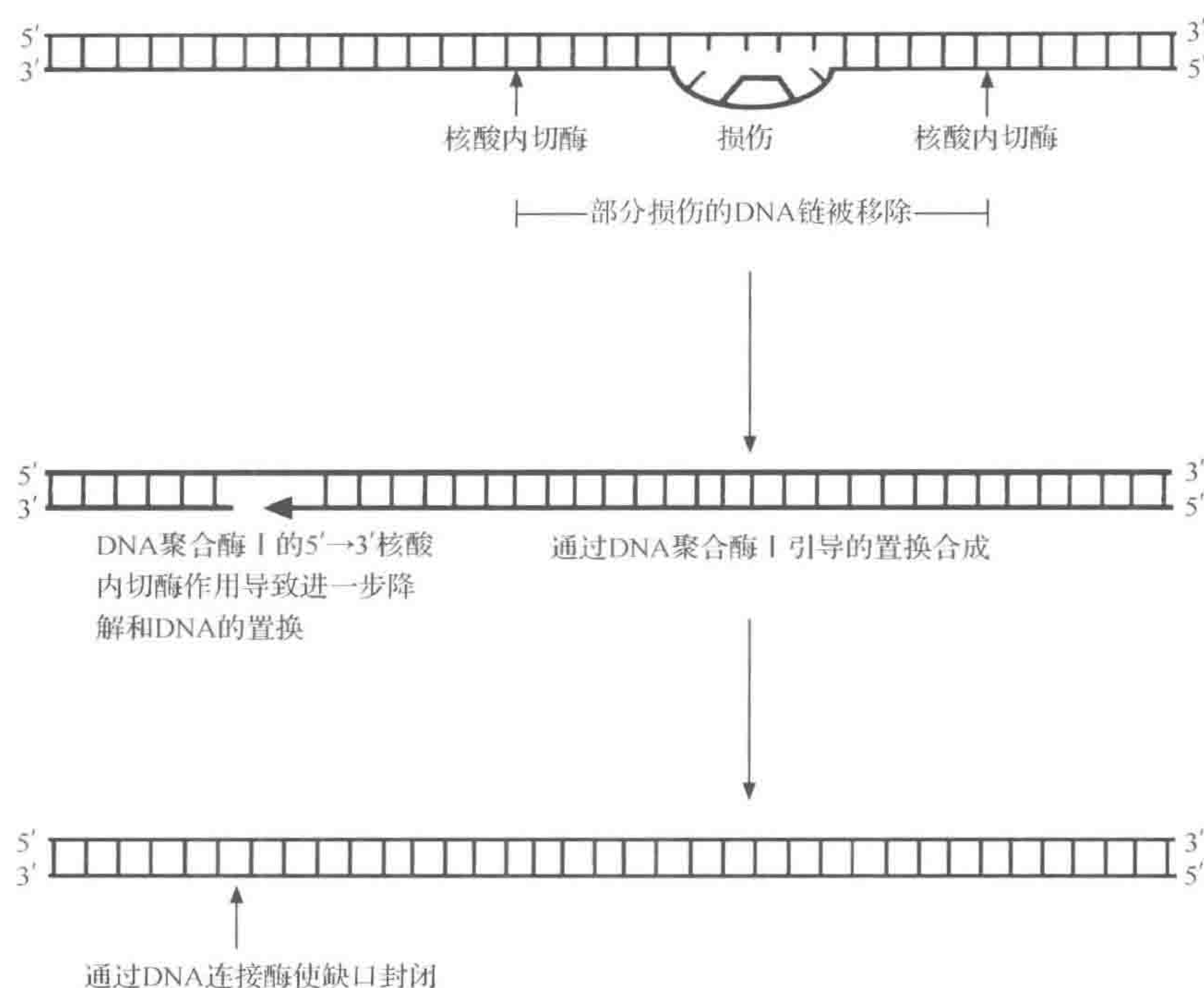


图 1.14 切除修复机制

核酸内切酶移除部分损伤 DNA 链，缺口由 DNA 聚合酶 Ⅰ 添补；DNA 聚合酶 Ⅰ 的 5′→3′ 核酸外切酶活性会移除更多的 DNA 并将其替代；最终，糖-磷酸盐骨架由 DNA 连接酶连接在一起

1.4.3 重组(复制后)修复

有另一种突变体，虽然具有切除修复功能，但对紫外线辐射异常敏感。这些细菌缺失负责基因重组因素之一的基因（*recA*）（见第 2 章）。这表明切除修复并不是处理紫外线损伤的唯一机制，还包括一般重组的一个进一步修复的机制。

干扰两条链碱基配对的 DNA 损伤形式通常会阻止复制，一部分原因是 DNA 聚合酶要求精确地配对 3′端，否则复制叉将因此暂停。但是，复制有可能越过损伤处重启，在新

合成链中留下一个空缺(图 1.15)。这部分 DNA 不能被一条完整链的切除修复所修复,然而,通过重组过程,用来自其他双链体的部分 DNA 可以填充空缺(即切断并重新连接 DNA)。尽管这仅是将损伤重新分类,而不是直接修复,但其确实达到了损伤时可修复的一种状态。最初的损伤可以通过切除修复进行修复,而其他 DNA 分子的空缺可以被 DNA 聚合酶Ⅰ和 DNA 连接酶填充。

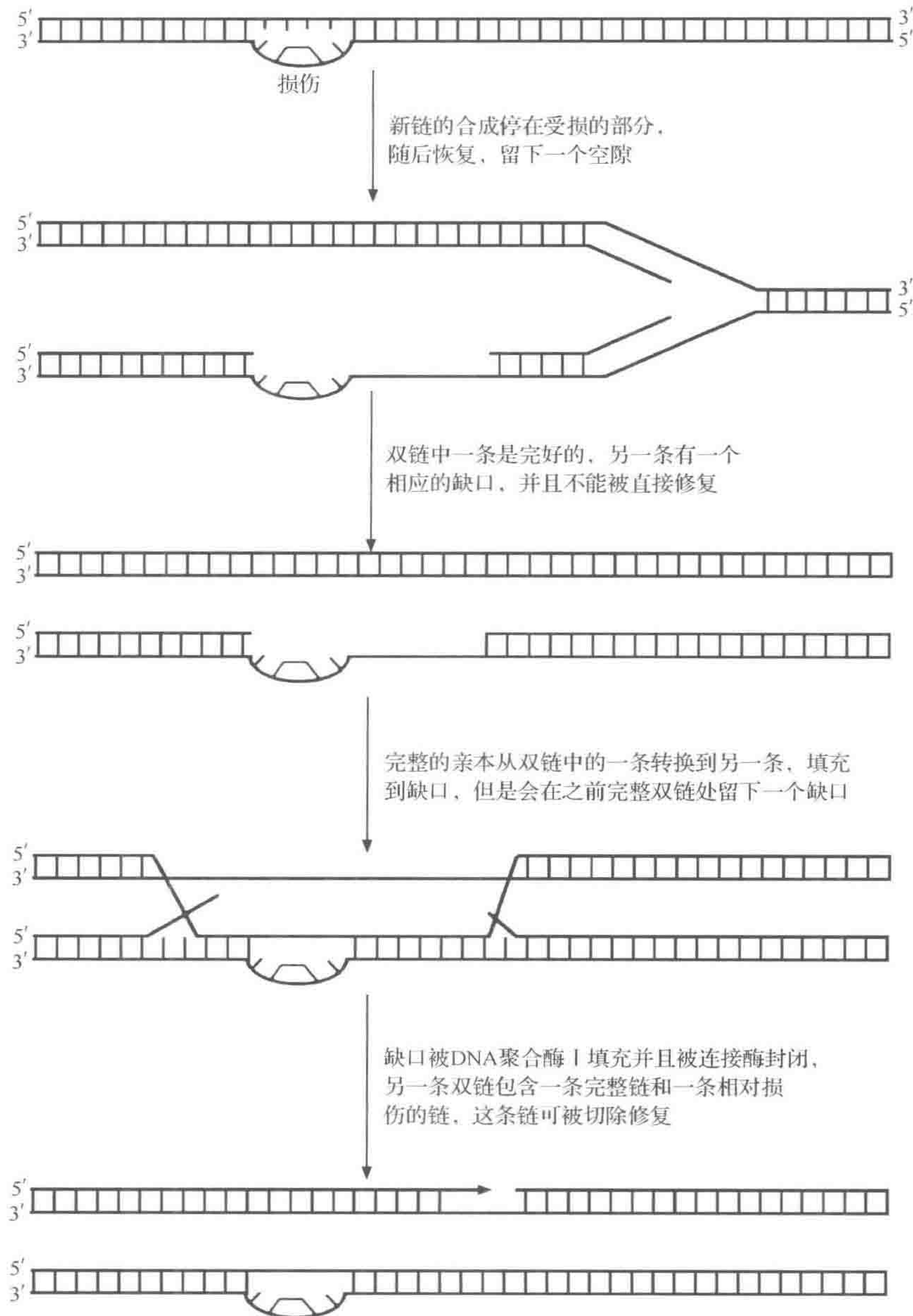


图 1.15 复制后修复

复制在损伤处终止,越过损伤处可重新开始复制,但在损伤处留下一段空隙。这一空隙可通过 DNA 交换进行修复,而最初的损伤通过切除机制进行修复

1.4.4 SOS 修复

当 DNA 受到严重损伤不能正常复制时，产生一种临时性修饰或诱导产生特定的 DNA 聚合酶，这使得其可以继续合成新的 DNA 链，尽管缺乏精确的 3′端。虽然生成了新的 DNA 链，但明显含有许多错误，故这个修复过程又被称为错误倾向修复。紫外线辐射产生的基因突变可以激发 SOS 修复，这在第 2 章涉及。

1.5 基因表达

遗传物质的表达主要是通过产生蛋白质来发生，涉及两个连续的将信息从一种形式转换成另一种形式的步骤：转录和翻译。不同的蛋白质组成不同的亚单位，每个部分都是由不同的 DNA 产生。完整的蛋白质是由几种不同基因产生的，任何基因的突变都可能导致其表达蛋白功能的缺失。顺反子即 DNA 分子中编码单一多肽链的区域，被用于强调单链和多聚体的区别。

1.5.1 转录

第一步是把信息转移到信使 RNA(mRNA)上。这个过程(转录)是在 RNA 聚合酶作用下完成的。像 DNA 合成一样，RNA 单链的合成方向也是 5′→3′。然而，转录和复制有显著的区别。首先，转录只产生相对较短的分子；其次，DNA 双螺旋中的一条链被转录(一些基因利用整条单链，一些基因利用其他的链，但总体来说，DNA 分子上任何特定区域都是从一条链上被转录的)，当转录形成一条单链后，它能利用单酶继续合成，无须合成后随链；再次，生成的相对较短的单链 RNA 导致较少的拓扑学问题，酶和 RNA 产物本来就能够使螺旋旋转，因此不需要解旋酶和拓扑异构酶参与复制；最后，RNA 聚合酶能够随时开始合成，不需要引物。因此转录比复制更简单。

启动子

由于转录导致合成相对较短的 mRNA 分子(通常长度为几千个碱基)，染色体周围有许多信号指导 RNA 聚合酶在正确的位置开始转录，并且当遇到终止基因时停止转录。另外一种说法是这种开始信号(启动子)能够传达关于转录的起始方向或者该以哪条链为模板的信号。

在大肠杆菌中，根据生长条件，在任何时候，RNA 聚合酶在合成 mRNA 时都为 2000～5000 个拷贝。RNA 聚合酶的基本结构是由 4 种多肽组成：两条完全相同的 α 链加上两条相关联但不完全相同的链(β 和 β′)。$\alpha_2\beta\beta'$结构为核心酶。启动子结合的特定位点是由于第 15 亚单位，σ(sigma)因子。包括 σ 因子的完整结构称作全酶。在第 3 章中，不同的 σ 因子识别不同的启动子，即基因的选择性表达。

当 RNA 聚合酶全酶结合在启动子区域后，这个初始结构(封闭型复合物)转变为两条 DNA 链间局部分开的开放型结构(图 1.16)。这揭示了编码链的本质，即用三磷酸核苷酸来合成 RNA。一磷酸核苷首先被合成，然后 σ 因子从复合物上分离，核心酶单独用

来延伸 RNA 链。新合成的 RNA 简短区域继续形成 DNA 模板,使得 DNA 链能够重新合成,因此允许 RNA 的继续合成,直到终止信号出现,mRNA 和 RNA 聚合酶被释放。

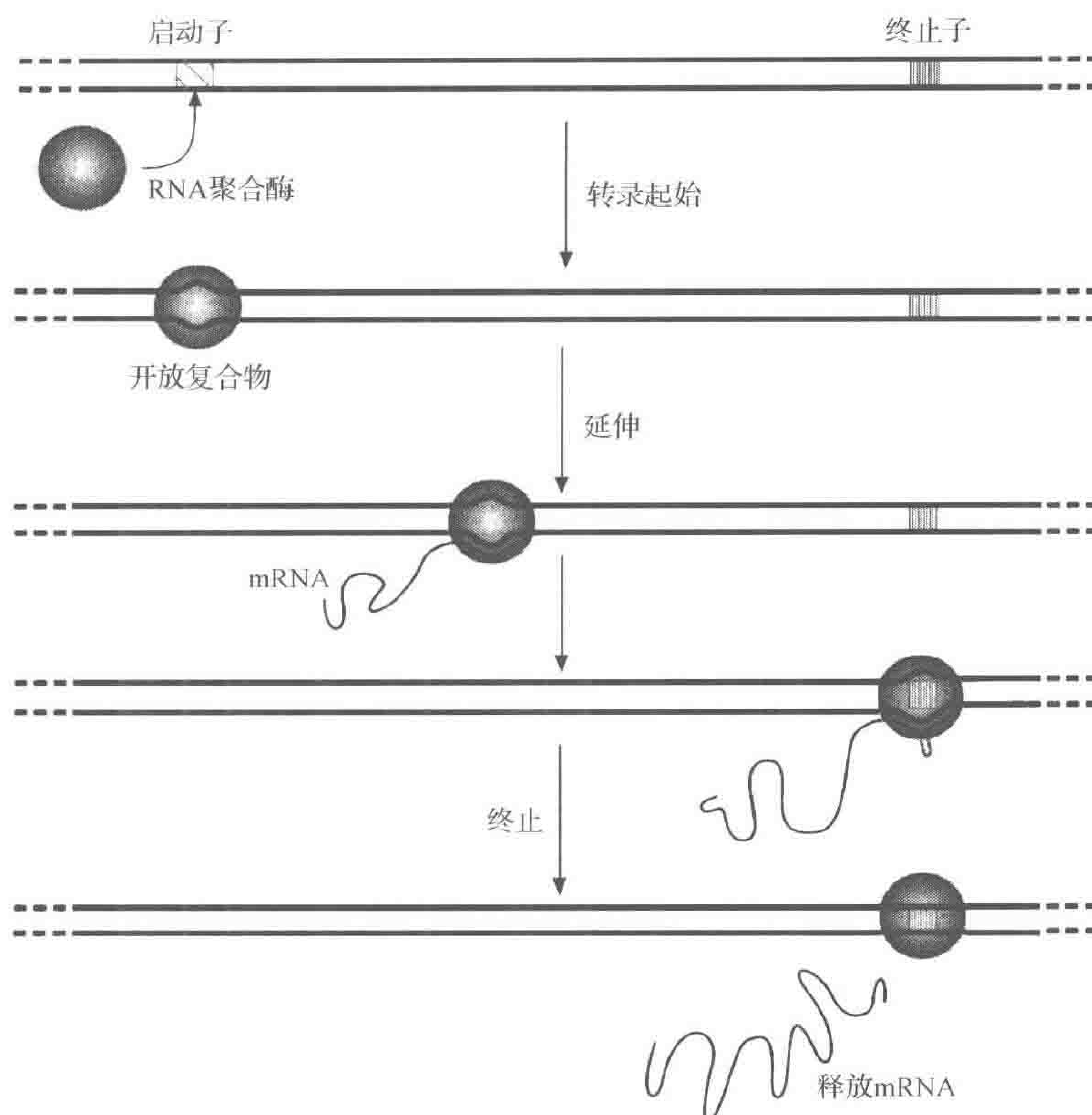


图 1.16　转录的主要机制

转录终止子

转录终止子的特征是一小段互补序列先于合成序列出现。当这个互补序列被转录,形成 RNA 时能产生一个茎-环结构,如图 1.17 所示。在大多数终止序列中,茎-环结构是由连续的尿苷酸残基终止的。

在图 1.18 中介绍了转录终止过程的模型。在双链 DNA 分开后的单链中,RNA 聚合酶需要一段大约 17bp 的没有损伤的 DNA。两条 DNA 链将会缠绕在一起,除非有什么东西阻止它们。

mRNA 本身就负责使 DNA 分子保持单链开放的状态,直到合成互补的 DNA 双链。在生理条件下,RNA-DNA 混合物比 DNA-DNA 配对的双链更稳定。DNA“鼓泡”大小是有限制的,通过拓扑异构来约束。螺旋结构解除在某种程度上使双链分开,从而产生分子卷曲,因此在两条链只能通过在两边增加缠绕的圈来分开。解除的区域越大,分子形成的压力越大。因此,mRNA 是游离于 DNA 模板之外的分子[图 1.18(a)]。

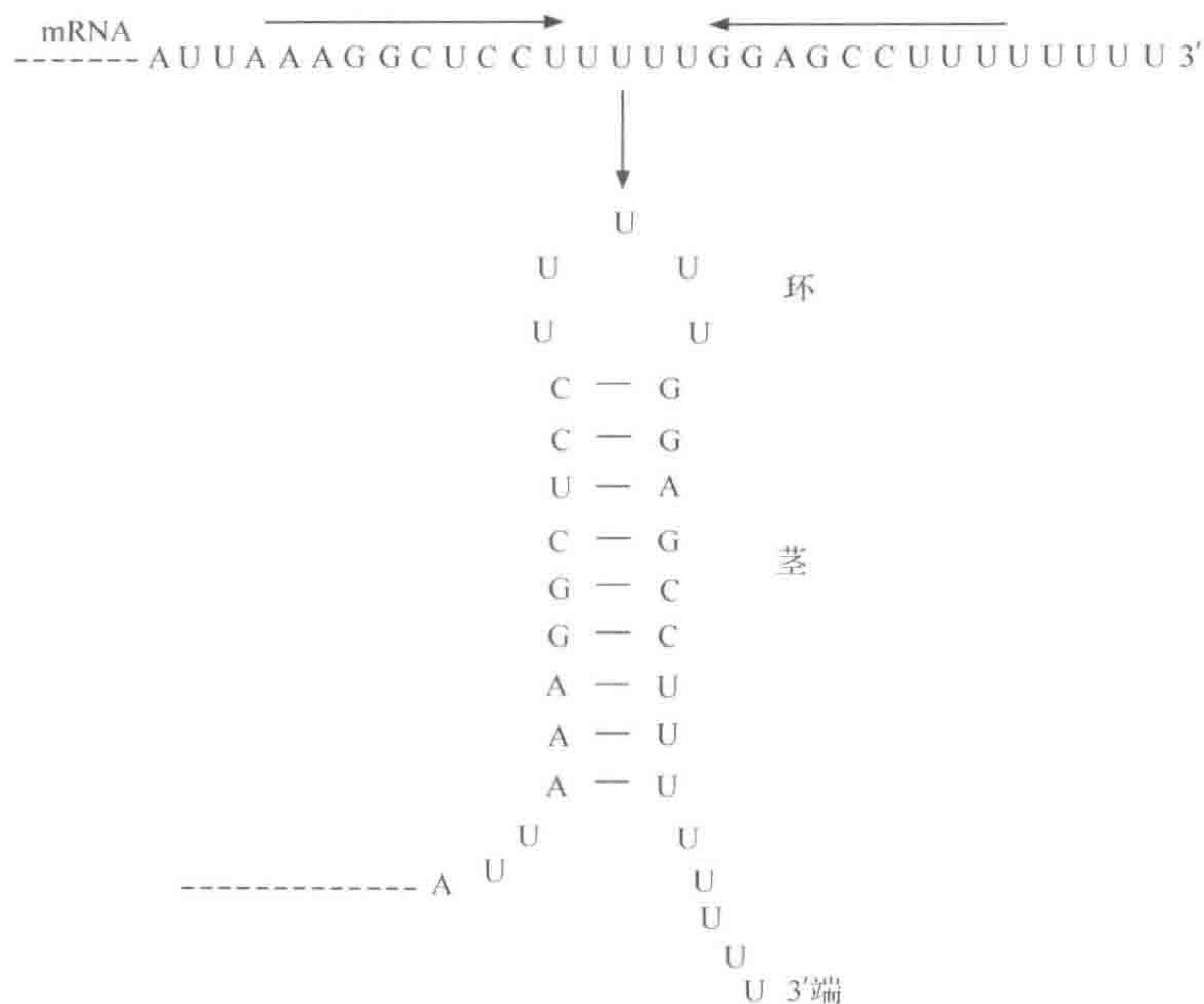


图 1.17 典型的终止子结构

箭头所示区域为互补区域，退火后形成茎-环结构

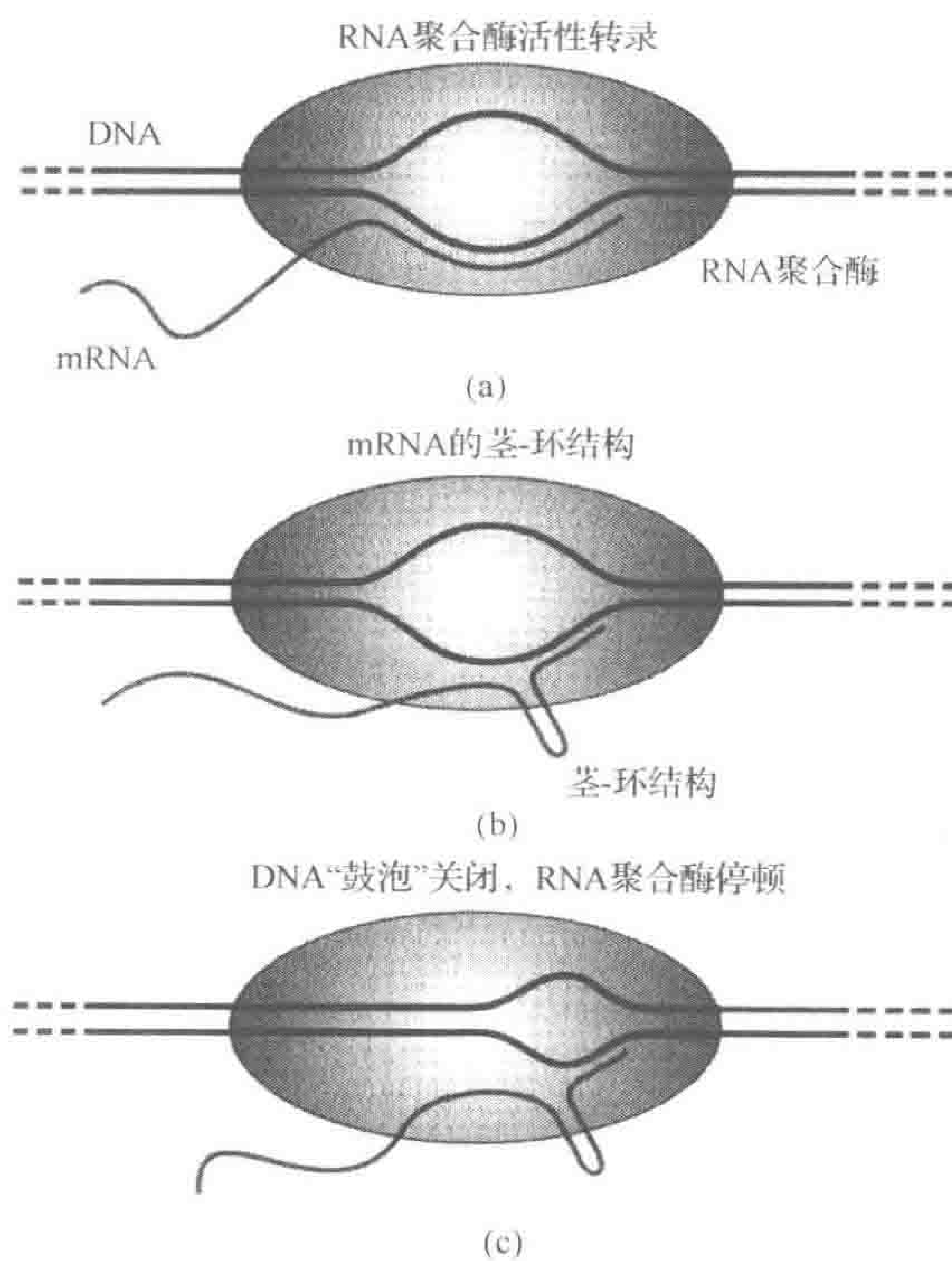


图 1.18 转录终止模型

(a) 转录时，mRNA 的存在使 DNA 一小段区域的双链打开；(b) 当到达转录终止位点时，mRNA 形成茎-环结构；(c) 这使 DNA 链重新配对，导致转录终止

当 RNA 聚合酶遇到转录终止子时会将其转录到 RNA 中，随后形成一个包含有部分 RNA 并能使 DNA 鼓泡保持开放状态的茎-环结构[图 1.18(b)]。鼓泡将会闭合，从而阻碍 RNA 聚合酶的活动。酶因此暂停发挥作用，一些碱基在茎-环结构的外面[图 1.18(c)]。由于典型的终止序列茎-环结构是以一串尿苷酸残基结尾(图 1.17)，mRNA 分子之间是由腺嘌呤和尿嘧啶相对较弱的氢键连接而成，因此，RNA 分子趋向游离于 DNA 模板，从而使 mRNA 合成终止。

一些终止子的茎-环结构不是以一串尿苷酸残基结尾。虽然 RNA 聚合酶可能在此位点暂停合成，但终止子依赖于另外一种蛋白[ρ(rho)因子]的活动，即依赖 ρ 的转录终止因子。

1.5.2 翻译

遗传密码子

mRNA 将携带信息通过基因编码的方式合成氨基酸序列(附录 F)，64 个核苷酸中每三个核苷酸(密码子)编码一种特定的氨基酸(或者是终止信号)。

密码子几乎是通用的，在不同的组织器官中，偶尔有小的差别，例如，UGA 的使用，在正常情况下是终止密码子，但也编码色氨酸和半胱氨酸。这说明所知的密码子一定是来源于早期的进化过程，后来固定下来，因为它能影响细胞中几乎每个基因的改变。

核糖体

细菌细胞核糖体由典型的两个亚单位组成，沉降系数分别为 50S 和 30S，整个结构被称为 70S 核糖体(沉降系数不是添加剂)。大亚基(50S)有两个 RNA 分子(23S 和 5S)外加 31 条不同的多肽链，然而小亚基(30S)含有一个 RNA 分子(16S)和 21 条多肽链。值得注意的是，真核细胞核糖体的结构和原核细胞在一些方面是不一样的。

核糖体 RNA(rRNA)分子通过大量的碱基对形成一个非常稳定的三维结构，与多种蛋白质结合形成核糖体，作为蛋白质生物合成的装配机。rRNA 的功能远不止这些，它还参与识别 mRNA(图 1.19)和催化肽键的形成。



图 1.19 核糖体结合位点(Shine-Dalgarno 序列即 SD 序列)

在细菌细胞中，核糖体附着在 mRNA 上的特定位置[核糖体结合位点，或者叫 RBS，也就是著名的 Shine-Dalgarno(SD)序列]，16S rRNA 的 3′端与其部分互补配对(图 1.19)，因此，核糖体的结合位点可以通过互补碱基配对间的氢键来调节。只要核糖

体结合位点空着就会与 mRNA 结合，因此，当 mRNA 还在合成时就已经开始翻译了(图 1.20)。

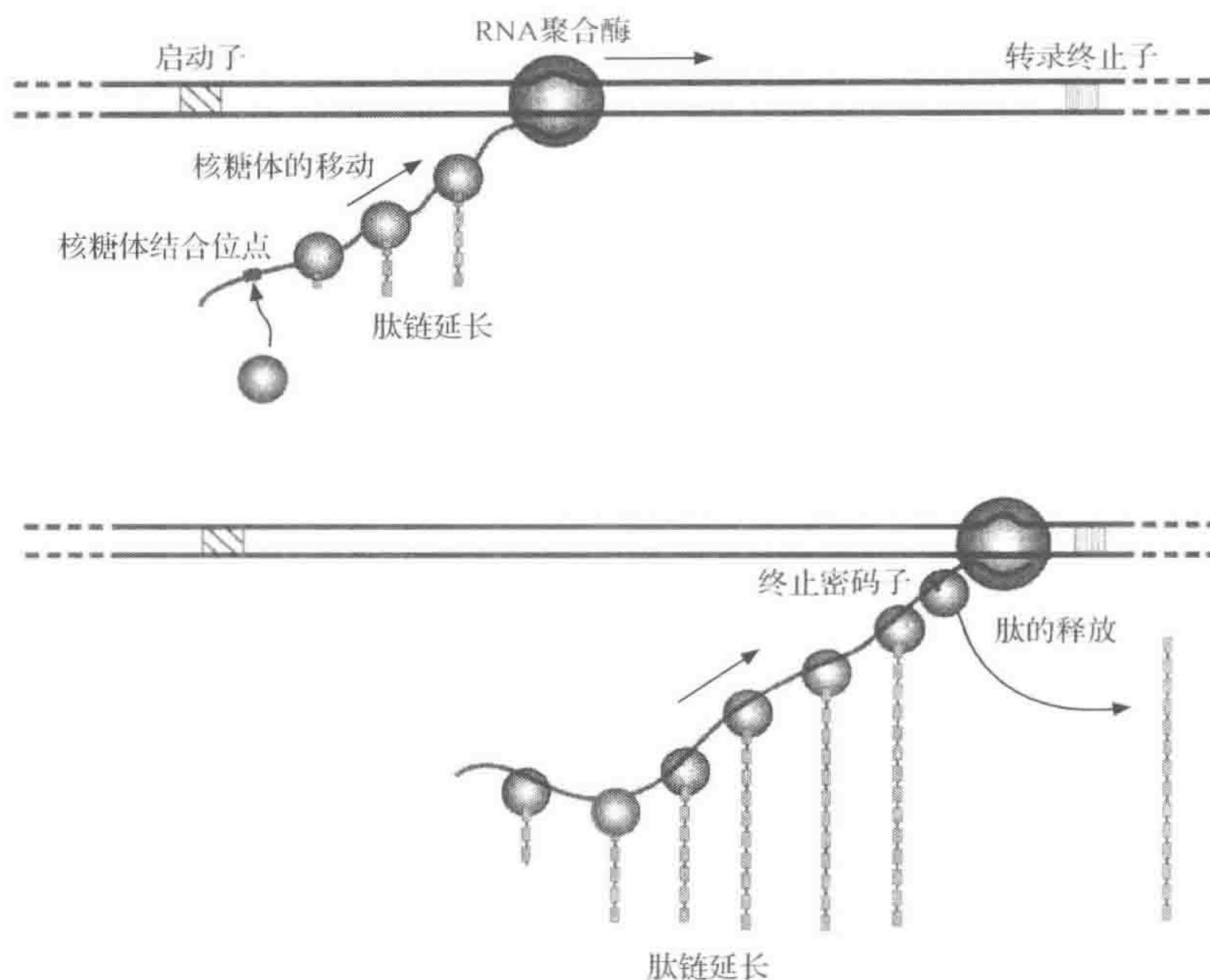


图 1.20　mRNA 的翻译

然而，并不是整个核糖体都参与 mRNA 翻译。不参与翻译的核糖体分离成 50S 和 30S 两个亚单位。翻译开始时，30S 亚单位结合在核糖体结合位点，tRNA 与相邻的起始密码子结合(通常是 AUG，但有时是 GUG，少数情况是 CUG)。50S 亚单位能附着在此起始复合物上，翻译过程就开始了。这些过程也需要一些非核糖体蛋白即起始因子的参与。

由于 mRNA 上是连续的相邻 3 个碱基一起被阅读(没有标点符号)，它能依据不同位置的起始密码编码 3 种完全不同的蛋白质，存在 3 个潜在的阅读框。核糖体结合蛋白与起始密码子的位置决定了阅读框。在第 2 章将讨论，一个碱基的增加和减少将会改变阅读框。

转移 RNA

密码子是由一种小分子 RNA 即转移 RNA(tRNA)识别。每种氨基酸至少有一种 tRNA 与之对应。然而，它们在结构上都很相似，包含着 75～100 个核苷酸的单链 RNA，折叠成三叶草结构(图 1.21)，实际的三维结构比简化的二维图形更复杂更紧凑。分子的两部分在合成蛋白质时有各自明确的功能：受体臂，由末端序列 5′端和 3′端配对的碱基所组成，接受活化的氨基酸(3′端被氨酰化)，反密码子臂含有三联体的反密码子，通过其与密码子配对而识别 mRNA 的遗传信息。

准确的氨基酸被一种特定的酶(氨酰 tRNA 合成酶)转移到 tRNA 上，这种酶具有很

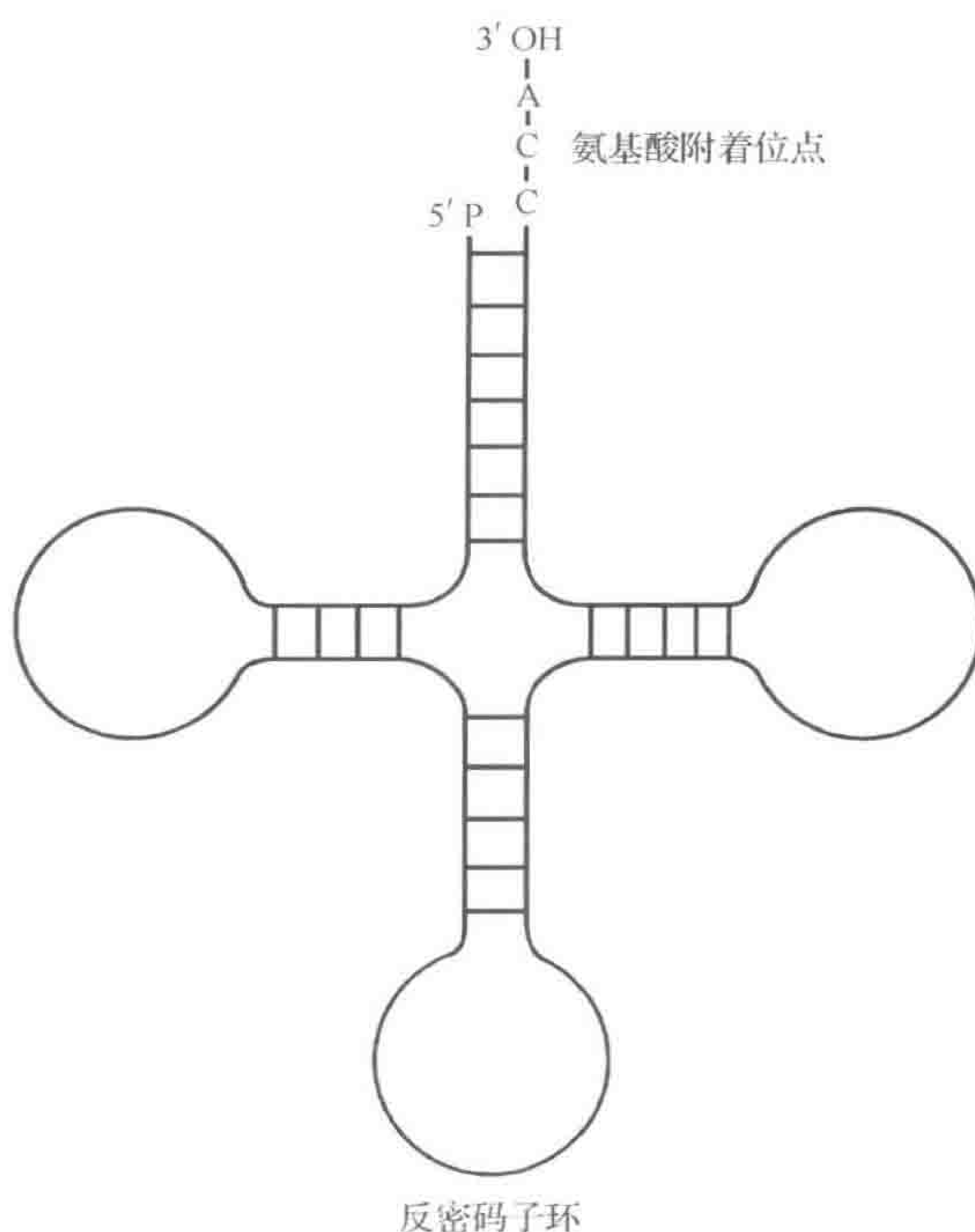


图 1.21 tRNA 结构示意图

特别的性质：一种酶能识别一种特定的 tRNA，并且识别与其对应的特定的氨基酸。例如，密码子 UGG（编码色氨酸）将被一个特定的 tRNA 识别，这个 tRNA 也将被色氨酰 tRNA 合成酶识别。这就保证了 tRNA 能够准确无误地转移相应的氨基酸。

因此，在特定过程中有 3 个独立的步骤：密码子-反密码子相互作用，通过氨酰 tRNA 合成酶识别特定的 tRNA，并且此氨酰 tRNA 合成酶识别相应的氨基酸。由于 tRNA 分子结构基本相似，并且一些氨基酸之间也十分相似，故偶尔发生错误不是令人吃惊的事情。低错误率（一种蛋白质分子中含有 1/1000 的错误氨基酸）是由于存在一种校正机制，即合成酶能够将错配在 tRNA 上的氨基酸释放出来。如果 3 个元素的特殊性是绝对的，那么将会至少有 61 种不同的 tRNA：每种对应一个密码子，但其中 3 个为终止密码子，即对应无意义的 tRNA。对多种氨基酸来说，的确存在多种 tRNA 对应不同特定密码子的情况。一些 tRNA 分子在细胞中的含量相对较低，说明在特定编码翻译过程中存在困难，这与一些特定的密码子频繁使用有关，这些密码子需要的 tRNA 种类很少，也很少出现。

然而，这并不完整。对许多 tRNA 分子来说，在密码子与反密码子的配对中并不是绝对的配对：密码子的第三个碱基在配对过程中有一定的自由度，可以摆动。一个相对复杂的规则（摆动假说）得以产生以解决密码子数与 tRNA 数不等的矛盾。因此，一些 tRNA 能够识别不止一个密码子。能识别全部密码子的 tRNA 数目少于 61（一般为 30～40）。

蛋白合成机制

在细菌中，起始密码子被特定的 tRNA 分子识别，即 $tRNA^{fMet}$。在起始密码子被 tRNA 识别之后，tRNA 分子被甲硫氨酸修饰，氨基酸被修饰成 *N*-甲酰甲硫氨酸。氨酰化的

tRNA 分子通常结合在核糖体上的 A(受体)位点，同时 tRNA 上的反密码子与 mRNA 上的密码子配对。只有当肽键形成之后，tRNA 才能够在核糖体上转移到第二个位点，即 P(多肽)位点。fMet-$tRNA^{fMet}$是唯一能够直接进入 P 位点的复合物。

$tRNA^{fMet}$反密码子能够识别 mRNA 上的起始密码子，在核糖体附近形成 30S 亚单位起始复合物。50S 核糖体亚单位也加入此起始复合物中(图 1.22)。经修饰过的 tRNA 与第二个密码子相结合进入核糖体上的 A 位点，fMet 残基转移到第二个氨基酸从而形成肽键。$tRNA^{fMet}$此刻未修饰而被释放，核糖体向 mRNA 3′端移动一个密码子，此时，仍与第二个密码子相结合的第二个 tRNA 分子从 A 位进入 P 位，即为移位。此时 A 位点空闲准备接受对应于第三个密码子修饰过的 tRNA。延伸由许多循环组成，每个循环包括肽键的形成、移位，结合另一个氨酰 tRNA 需要一些其他的非核糖体蛋白(延伸因子)的参与以及 GTP 水解释放的能量。

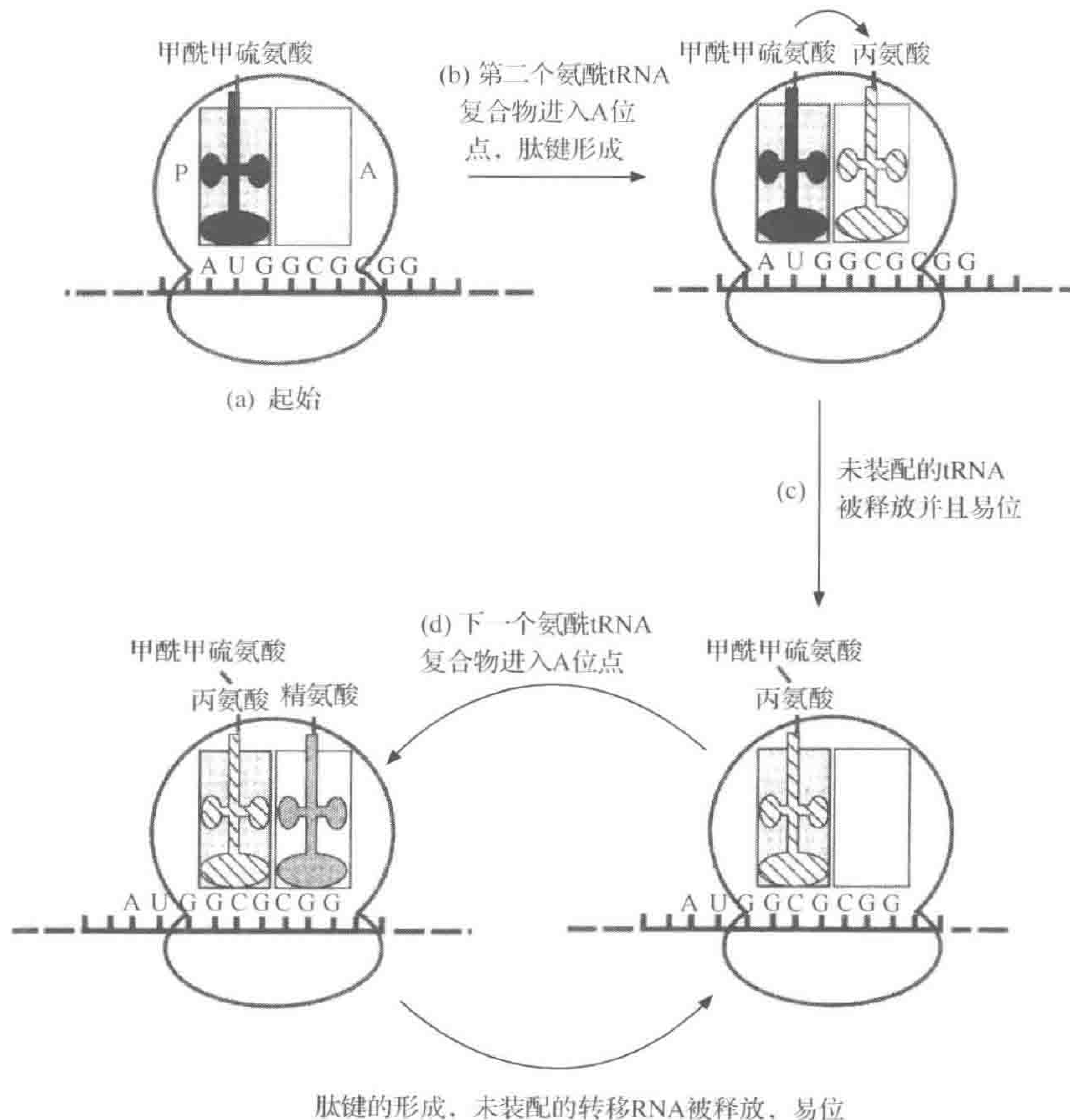


图 1.22 蛋白质合成机制略图

(a) 核糖体 30S 小亚基结合到 mRNA 上，形成最初的起始复合物，随后 50S 大亚基结合到起始复合物上；(b) 第二个 tRNA 进入 A 位，通过将 N 端的甲酰甲硫氨酸(fMet)转移给第二个氨基酸的游离氨基而形成第一个肽键；(c) 无负载的第一个 tRNA 被释放，核糖体沿 mRNA 移动，这时第二个 tRNA 进入 P 位，A 位被空出以接受下一个 tRNA；(d) 下一个 tRNA 进入 A 位，如(b)所示进行下一肽键的生成；重复进行(c)和(d)，直至到达终止子

当核糖体移动到足够远时，mRNA 分子上的核糖体结合蛋白再次暴露出来以便其他的核糖体能够附着。mRNA 的单链分子上携带很多个核糖体参与序列的翻译(见图 1.20)。

每个核糖体合成蛋白质的方向是沿着 mRNA 5′→3′移动直到终止密码子出现。由于没有任何氨酰 tRNA 与此终止密码子结合，因此不能把此密码子翻译成氨基酸，翻译终止。多肽链和释放因子被释放出来，核糖体从 mRNA 上脱离下来。

1.5.3 翻译后事件

翻译形成的多肽链必须经过进一步的修饰后才具有生物功能。生物信息学功能的激活涉及几个步骤。蛋白质必须正确地折叠。除了一级结构(氨基酸的顺序)之外，有 3 种构象水平。二级结构是指氨基酸的空间排列，通常形成的结构为 α 螺旋和 β 折叠(图 1.23)，分子间由非共价键(如氢键)来稳定。蛋白质的不同区域具有不同的二级结构。

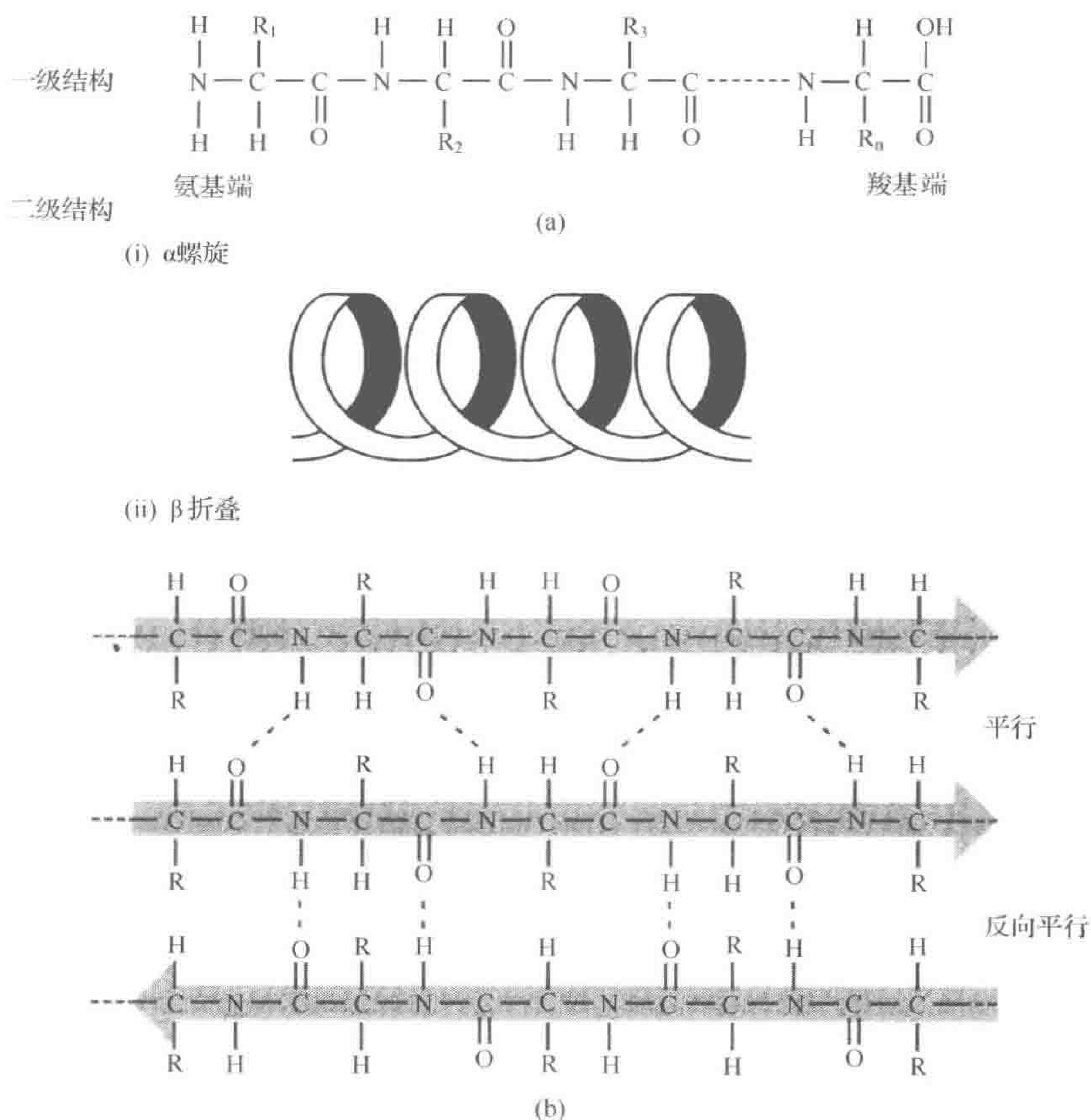


图 1.23 蛋白质的一级和二级结构

(a) 蛋白质的一级结构即氨基酸的排列顺序；(b) 两个主要的蛋白质二级结构，α 螺旋(图 i)和 β 折叠(图 ii)

二级结构进一步折叠形成三级结构(图 1.24)。分子间由非共价键以及半胱氨酸残基间的疏水键来稳定。三级结构可能包括两个或者更多区域即结构域。许多蛋白质是由

几条多肽链组成，蛋白质分子中各亚基的空间排布及亚基接触部位的布局和相互作用，称为蛋白质的四级结构(图 1.25)。

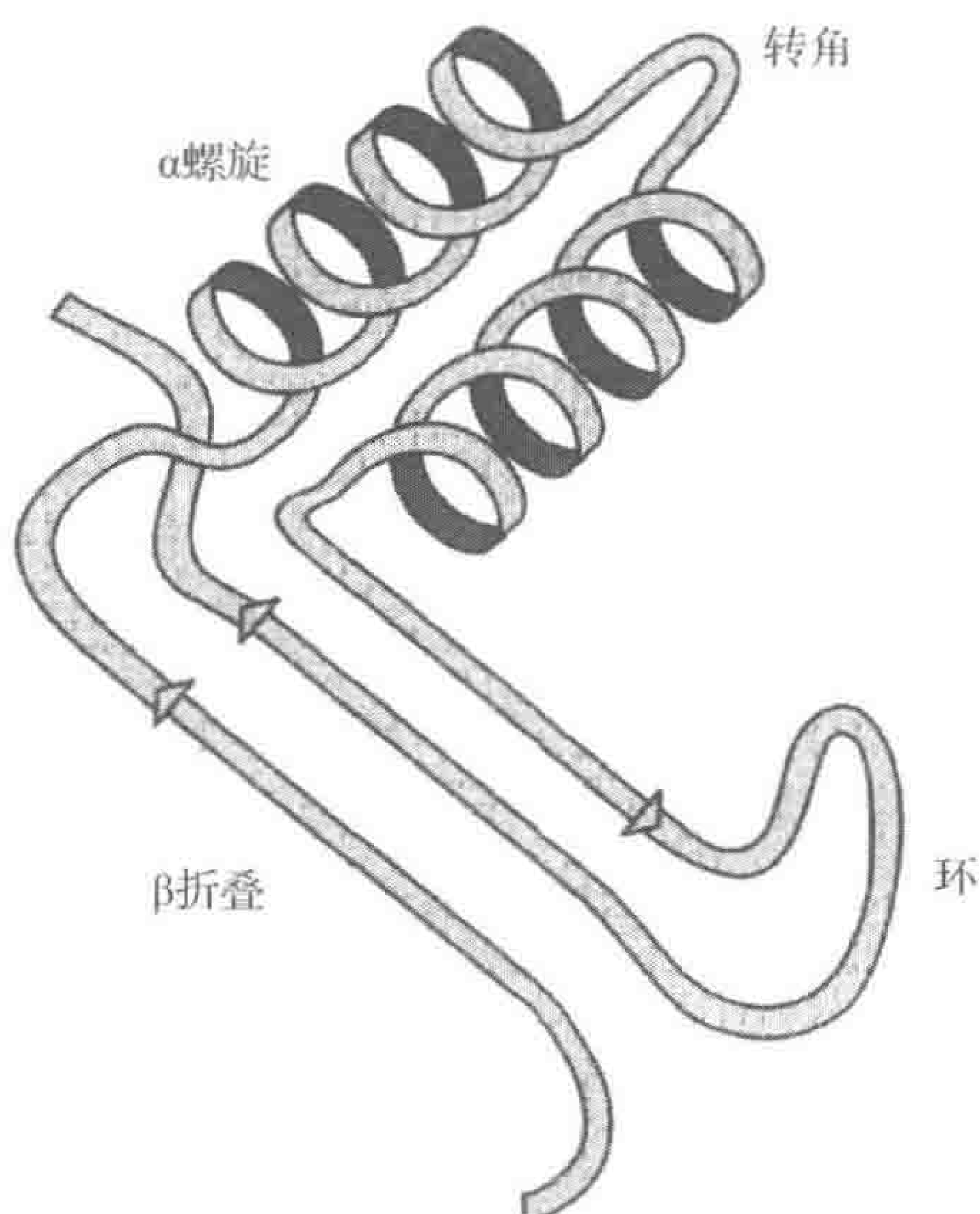


图 1.24　蛋白质的二级和三级结构

蛋白质的三级结构是指二级结构单元的排列方式，本图所示三级结构为两个中间由转角相连的 α 螺旋，一个 β 折叠，以及相对松散的不规则区域，即由一个环组成

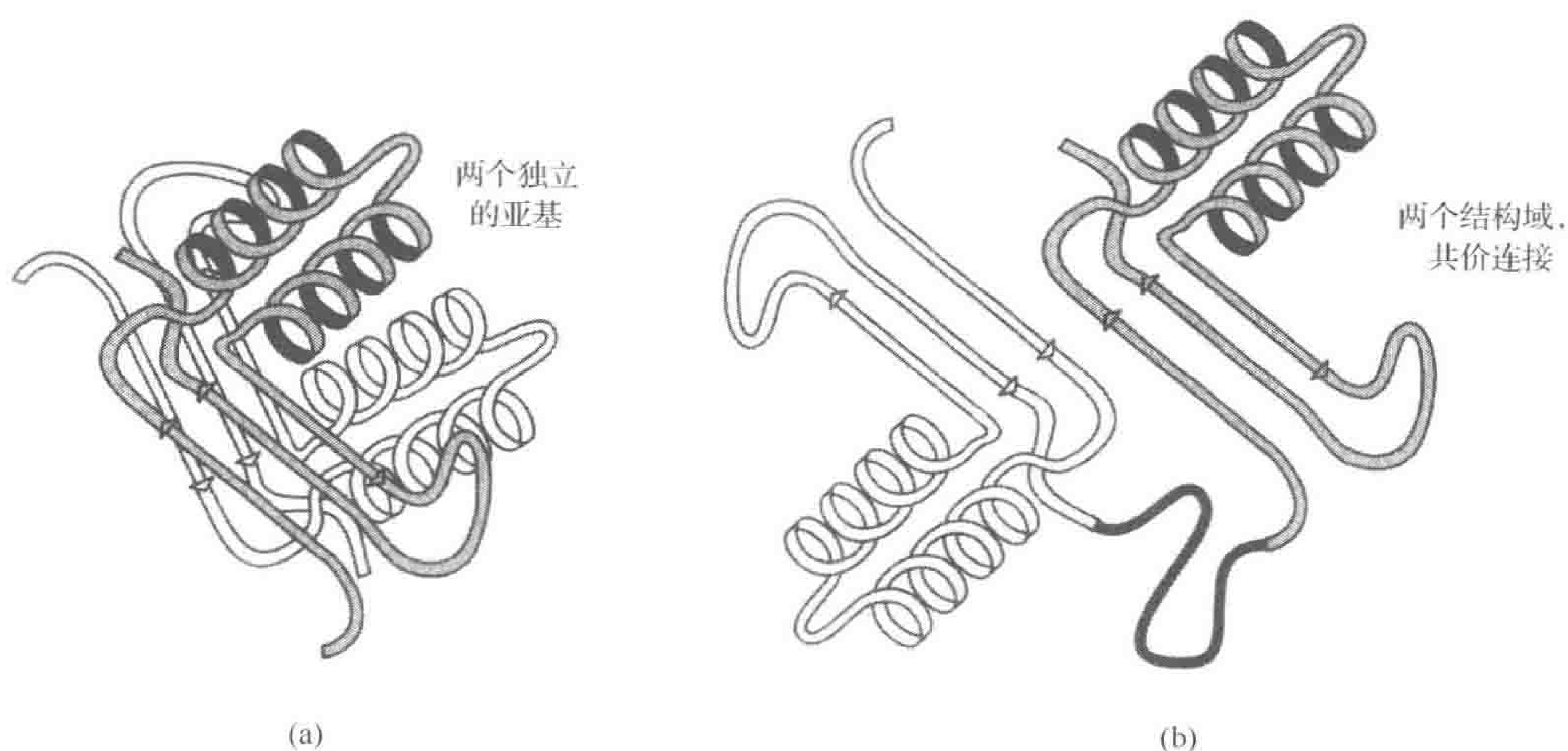


图 1.25　蛋白质结构：亚基和结构域

(a) 一个蛋白质可能由两个或两个以上独立的多肽链(相同或不同)组成，这种亚基间的结合方式称为四级结构；(b) 单一的多肽链可以折叠成几个独立的结构域，这些结构域分别具有不同的功能，一种来源的酶可能具有两个分离的亚基，而不同来源的相关的酶可能是含有两个结构域的单一肽链

很难预测多肽链的空间结构，一部分原因是可能性太多，另外一部分原因是多肽链将形成二级(更高级)结构，因此，在结构上的约束是由肽链的后部分决定的。最终结构是经

过整个多肽链重新折叠后所形成的热力学上的稳定结构。

此外,多肽链的折叠并非完全自发。细胞含有分子伴侣蛋白,从而使蛋白质获得正确的构象,例如,新生多肽的相互作用可以阻止其采用错误的构象形式直到产生完整的蛋白质。分子伴侣在蛋白质变性过程中也具有重要的作用,它能够提供抵抗热和其他压力状态的保护限度。一些分子伴侣是在热变性蛋白累积到一定程度后产生的,即热休克蛋白或者应激蛋白。分子伴侣另外的功能是收集多肽亚单位形成多聚蛋白或者更大的结构,例如,噬菌体头部的聚集(第 4 章)可能就需要分子伴侣的参与。

分泌

许多细菌蛋白具有分泌到细胞表面或者细胞外环境中的功能。蛋白质分泌出来的第一道路障即细胞膜,最普通的蛋白质跨膜转运机制即一般分泌途径(GSP,有时也称为分泌依赖性途径)。所有蛋白质利用这个系统在 N 端具有一个特定序列,使目标蛋白经过此分泌途径并在转运完成后分开。在大多数革兰氏阳性菌中,这种分泌机制能够使蛋白质分泌到细胞表面或者周围介质。然而,革兰氏阴性菌有层外膜并且 GSP 本身就能够运输,但不能运到细胞外,而是运到周质中,即细胞质和外膜之间的区域。

尽管在蛋白质分泌方式上革兰氏阴性菌没有阳性菌那么丰富,但是它们有重要的分泌机制。最常见的为Ⅱ型机制,它依赖于分泌依赖性途径将蛋白质转运至周质,然后利用特殊的多种蛋白质复合体将蛋白质转运过外膜。

大多数革兰氏阴性菌(和一些革兰氏阳性菌)也拥有一些独立于 GSP 的分泌系统,即为非分泌依赖性途径。因为Ⅱ型分泌系统和在第 6 章中描述的结合质粒具有许多相似的特性,可以被细菌用于将蛋白质介导进入真核细胞,这是值得注意的。也许最著名的分泌系统是Ⅲ型途径,它被比喻成分子注射器,并且可被用于直接把蛋白质注入真核细胞的细胞质中(这种系统在第 9 章会详细描述)。分泌是在细菌和宿主细胞直接接触的情况下发生的。结果,这种途径被病原体广泛应用于向宿主细胞介导效应分子,因此,对细胞正常功能的破坏就有利于细菌。

其他翻译后修饰

除了上述翻译后修饰外,蛋白质可能要经过一系列其他的翻译后修饰,如糖基化、生物素化、添加脂质和蛋白质裂解。这一系列的修饰太复杂而不能在此全部描述,但结果是最终的蛋白质结构在很大程度上由细胞本身决定。由于翻译后修饰是多肽链获得生物学功能所必需的步骤,很难获得精确的翻译后修饰过程,它能够影响功能基因在异源宿主中的表达结果(见第 8 章)。

1.6 基因组织

在细菌中,功能相关的基因常(不总是)处于同一组即操纵子中(图 1.26)。一个操纵子有一个启动子,启动子被转录成一条多顺反子 mRNA 分子,并且它携带着几个蛋白质的信息。这一组基因相互协调:细胞在适当条件下生长将导致操纵子中的所有基因同时相互协调工作。这部分将在第 3 章进行全面讨论。

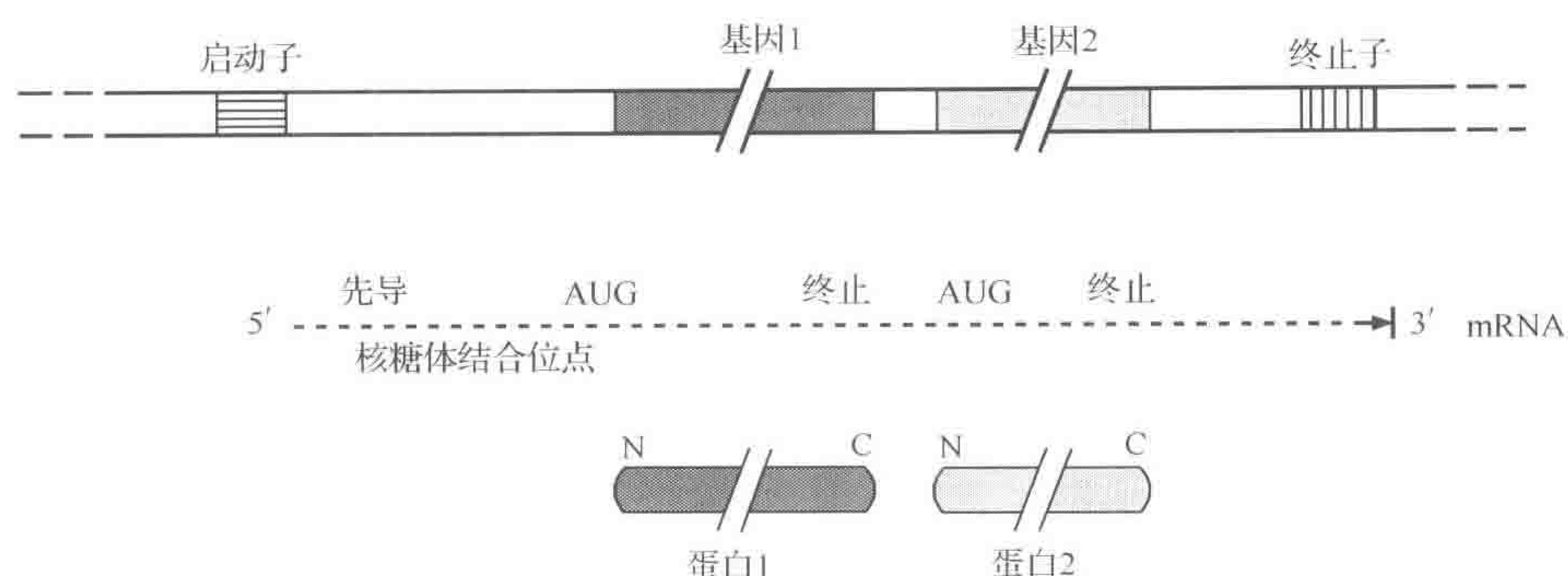


图 1.26 细菌操纵子的结构和表达

一个典型的操纵子有一个启动子，可转录出一个多顺反子 mRNA，该 mRNA 可翻译出多条独立的多肽链

当核糖体翻译操纵子中的第一个顺反子之后，它将和 mRNA 分开，因为翻译下一个顺反子需要核糖体附着在相邻的第二个顺反子的起始密码子上。在一些情况下，第二个基因的起始密码子和第一个基因的终止密码子隔得非常近(实际上序列部分重叠)。如果是这种情况，当第一条多肽链释放之后，核糖体在邻近的起始密码子上再次开始翻译，不会与 mRNA 分开。

原核生物和真核生物之间的一些主要区别值得注意。真核生物中的 mRNA 扮演着真正信使的角色，mRNA 在细胞核内产生，转移到细胞质的过程中即发生翻译。在原核生物中，转录和翻译几乎同时发生，核糖体一发现 mRNA 上的核糖体结合蛋白就立即附着在上面。因此，在原核生物中，当 mRNA 被合成后随即参与翻译。大多数原核生物 mRNA 的存在极短暂，半衰期只有几分钟，这比 mRNA 本身合成时间以及参与翻译所需时间都短，这就导致了转录和翻译同时发生。原核和真核之间的另外一个区别在于核糖体结合位点，真核生物中 mRNA(通常)只编码一条多肽链。真核细胞的核糖体结合在 mRNA 上的 5′端，然后向着 3′端移动直到找到起始密码子为止(在真核细胞，确实存在内部核糖体进入位点，但这只是例外，而非普遍。)。

在真核细胞中，转录的第一个产物是前体 mRNA。这种前体 mRNA 只存在于细胞核，它包含的附加序列(内含子)在加工修饰后被去除。在一些情况下，最终 mRNA 的大小要比原始基因小 10%。通常来说，原核生物基因中没有内含子，但是有少数原核基因(主要是噬菌体)中确实含有内含子。

最后，真核生物 mRNA 通常(不总是)多腺苷酸化，3′端有一串腺嘌呤残基。多腺苷酸的尾巴结构通常被用于纯化真核细胞 mRNA。原核细胞 mRNA 不含连续的多腺苷酸，尽管小部分原核细胞 RNA 分子携带着短寡核苷酸尾巴。

2 突变与变异

在基因操作时代来临之前，相关基因发生改变（通常是有害的）的细菌特异突变株的分离是细菌遗传学的基石。这种改变表现为生物体相关特征（表型）的变化。人们能够用生物体的这种可以观察到的变化跟踪基因的传递。根据生物体的可见特征能够推导其遗传本质（基因型）。

由于基因克隆技术不仅使人们可以在之前没有获得突变体的情况下分离和研究特异基因的结构，而且对该基因结构进行体外操作，因此，该技术为遗传学研究提供了全新的视野。在本书后边章节将领略到基因克隆技术所产生的巨大影响，但对于和细胞代谢有关的基因功能的了解则要应用体内分析来完成。

任何细菌群都是不均一的。大肠杆菌的培养物（在最佳条件下）用 10～15h，约经 30 个世代，由单个细胞生长至最高浓度（通常约为 10^9 个细胞/mL）。在培养物中，由于生理因素或遗传的改变，会出现从一个细胞变为另一个细胞的变异。生理变异指的是在任一时刻，由于环境、生活史或生长阶段的不同，细胞对外部影响产生不同的反应。生理改变和遗传变异的关键差别在于发生改变的特征能否遗传。

2.1 变异与进化

现代进化理论源于查尔斯・达尔文（Charles Darwin）在《物种起源》（1859 年）中提出的和阿尔弗雷德・拉塞尔・华莱士（Alfred Russel Wallace）提出的概念，即物种通过自然选择过程发生进化，更适应环境的个体得以生存、繁衍，并将其特征传递给后代（框 2.1）。当时，该理论存在两个不足之处：第一，对遗传机制的理解尚不清楚；第二，对种群变异的本质没有给出令人满意的解释。

框 2.1　进化、神创论及天才设计

遗传学的科学研究是以进化论为基础的。但有些人坚持除了进化，物种起源还应包括其他观点（或在极端情况取代进化），即物种是神的创造（神创论）或天才设计（神创论的不同名称），也就是一位天才创造者创造了活物。

这些批评家宣称进化“仅仅是一种理论”，这反映了他们对“理论”这一概念认识的局限性。牛津英语大词典对“理论”一词有几种解释，包括“推测（尤指臆造）的观点（某人最偏爱的观点）”和“摘要性的知识或推测的想法（理论上都很好）”。这是日常用语中常用的含义，但在科学领域中，则用“假说”表达上述含义。牛津英语大词典中对“理论”

还有一种解释:"不依赖于被解释的某一特定事例,而是基于具有普遍性的原则解释某些事情的推测或系统的推论(如原子理论、进化论)"。

因此,进化论不是推测性的想法,是基于观察获得的许多想法的综合,以解释物种是如何起源(并继续进化)的。进化论可以被概括为以下几点。

(1) 在一个物种中,变异自发产生。

(2) 某些变异体能更好(或不是很好)地适应一种特殊的环境。

(3) 更好地适应环境的变异体将更好地生长、繁衍,生存下去。

(4) 如果同一物种分为两个群体(如地理因素),为更好地适应各自的环境,它们将逐渐分化。

(5) 最终,两个群体的差别足以将其称为不同的物种。

进化论也在不断地完善中,如解释亚种如何分离。两种不能很好异种交配的变异体(如在不同时间开花的植物变异体)将进一步分离成两个不同的亚种。

很难想象一个复杂的结构,如眼睛能够进化产生,就像毫无根据地假设这种处于非完美状态的机体是毫无用处的一样。事实上,在自然界中,有着大量不同的光敏结构,从一些最简单的单细胞趋光生物体到具有明显优势的进行光合作用的生物体。所有这些光敏结构都需要受到光源刺激时将信号传递给细胞其他结构的,使之向光源移动的光敏感膜蛋白。基于这样一个简单的开始,不难设想,假设存在大量的生物体,经历了几百万年的进化发展到今天,便在不同物种中发育成为多种多样的"眼睛"。

关于第一点,乔治·孟德尔(Gregor Mendel)根据他的研究结果提出:遗传是分离单位的组合,而非父母特征的混合。这一结论提出于 1865 年,但一直被人们所忽视,直至 1900 年,人们对(真核细胞)染色体功能的认识的加深,为理解孟德尔的研究结果奠定了基础。

对于变异的本质,尽管人们随后普遍接受了变异理论,即变异是不受环境影响的随机发生的过程。但达尔文在这一方面的观点与早期由拉马克提出并统治了人们思想 50 年的获得性遗传观点相似(举一个滑稽的例子,长颈鹿因为要不断伸长脖子够树枝上的叶子,所以脖子长得很长)。20 世纪早期,不断积累的证据推翻了拉马克的观点,人们认同了随机变异。但在 1964 年以前,苏联人一直坚持拉马克主义,李森科(Lysenko,20 世纪生物科学史研究中的著名人物之一,苏联生物学家、农学家,在苏联斯大林时代具有极大的影响力)相信环境变化能引起永久性的遗传变化,他的权威性使苏联的遗传学界一直受这一思想控制。

因此,尽管进化观点常被称为达尔文主义,但实际上它是源于多种思想的综合体。

2.1.1 彷徨变异实验

当将细菌暴露于选择性条件时,会产生某种类型的突变,由此看来,某些条件确实能诱导相关突变,因此微生物学家在相当长的时期内一直坚持拉马克主义,直到很多遗传学家提出质疑,情况才有所改变。1943 年,Luria 和 Delbruck 发表了他们的研究结果,并成

为拉马克学说灭亡的标志。

他们检测了具有抗噬菌体 T1 抗性的大肠杆菌突变体的出现。通过将细菌培养物简单暴露于噬菌体,可以很容易地分离到这样的突变体。问题在于:变异体是在细菌暴露于噬菌体之前就已经出现(随机变异)的,还是暴露于噬菌体之后(并且仅仅是因为暴露于噬菌体的缘故)才出现的(定向变异)。

为了回答这个问题,他们设计了如下实验(图 2.1)。在一系列细菌试管中分别接种少量噬菌体,这些噬菌体取自同一管培养物,培养一定时间后,计算每一试管中抗噬菌体大肠杆菌的数量,如果是由于加入噬菌体才产生的突变,每管中出现的突变株数量应接近,结果服从统计及取样方差。而且不管是从单独培养的试管里取样,还是从一系列独立培养的小试管里取样,突变株的数量及变异量应该相同。

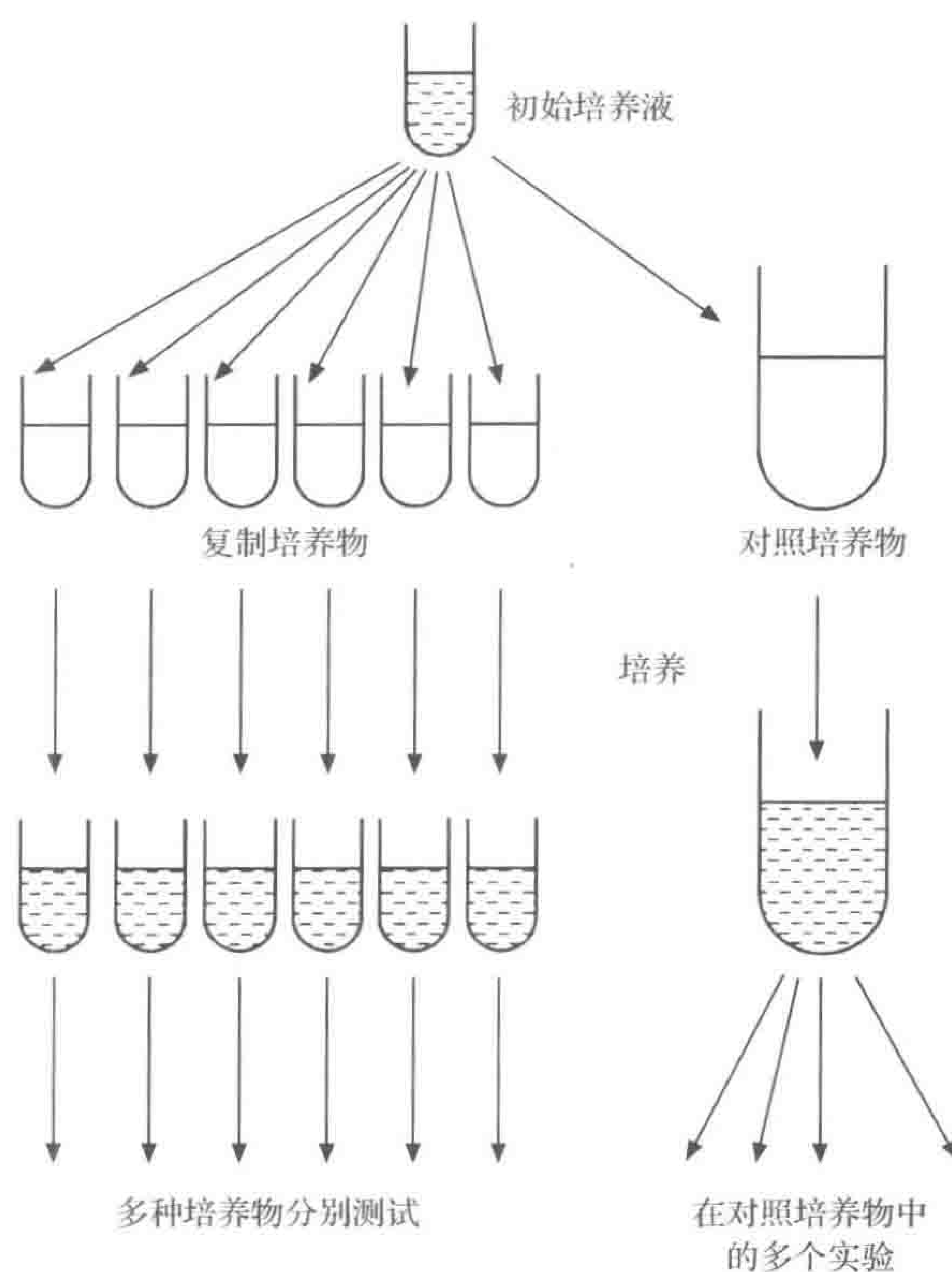


图 2.1 彷徨变异实验

实验中设计一系列重复培养物以检测特定的突变体,如果突变随机发生,每个重复培养物中的突变体数量要远高于取自单一试管的对照培养物中的突变体数量(图 2.2)

如果突变是暴露在噬菌体前随机发生的,因为突变可以发生在培养过程中的任一时间,那么不同试管内产生的抗噬菌体突变体数量差异是相当大的。如图 2.2 所示,如果突变发生在较晚的阶段,则产生的突变比例较小;如果突变发生在早期,则将产生大量的突变体。获得的实验结果为,一系列独立培养的细菌间发生的变异远远高于在一个试管里取样的细菌间的变异,支持了随机变异假说。

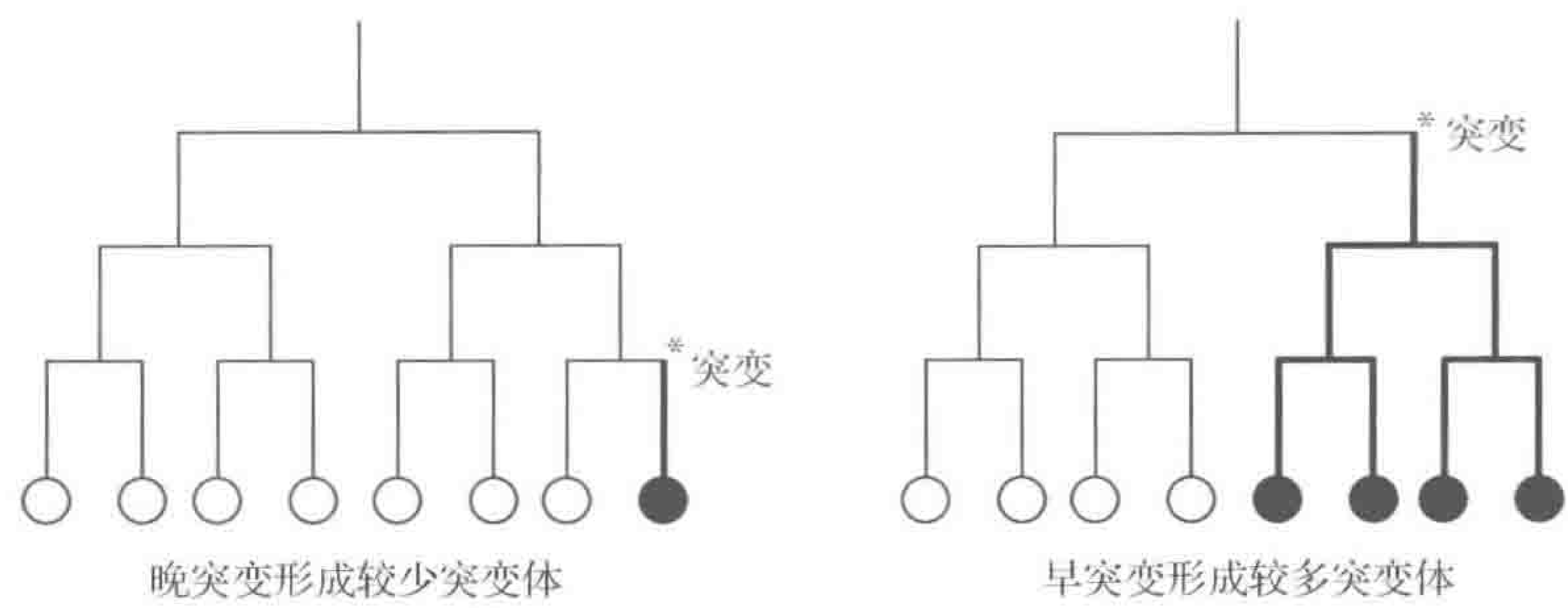


图 2.2 彷徨变异实验原理

随机突变可发生于细胞生长的任一时间，如果突变发生于晚期，培养物中突变体比例较低，如果发生于早期，突变体比例则大大增加

2.1.2 影印平板法

影印平板法更直接地验证了变异是随机产生的。具体过程如下(图 2.3)：将一块灭菌丝绒垫在一个长有许多细菌克隆的琼脂平板表面轻轻压一下，每一克隆都有少量转移到丝绒垫上，然后将这些细菌接种到一个或多个新平板上。经培养，这些平板上都会长出作为母平板复制物的一系列细菌克隆。如果将其中一个平板暴露于 T1 噬菌体，敏感的细菌克隆死亡，抗性克隆将存活。在未暴露于噬菌体的平板上挑取对应的细菌克隆，即使

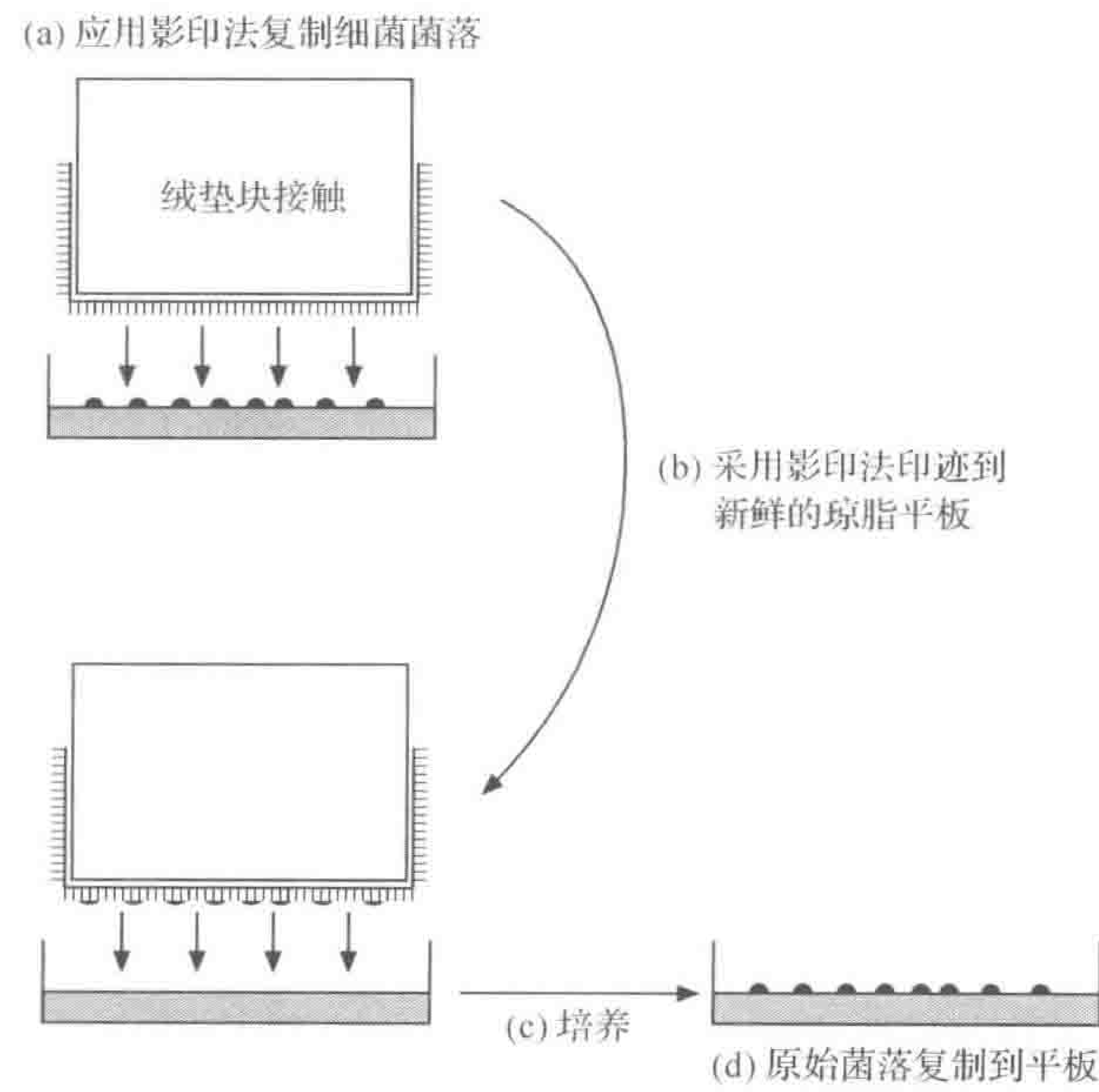


图 2.3 影印平板法

(a) 将包裹在木块上的一块灭菌丝绒轻压于生长有细菌克隆的平板表面，每一克隆都有少量转移到丝绒上；(b) 将此丝绒在一新平板上印一下，以培养母平板的影印平板；(c) 经培养，除了那些不能在新培养基上生长的细菌克隆，在新平板上长出与母平板一致的细菌克隆，而那些不能生长的克隆可在母平板上获得(d)

从未接触过 T1，它们对 T1 也具有抵抗力。这就证明，具有抵抗力的菌株是随机变异产生的，而非暴露于噬菌体的结果。

2.1.3　细菌的定向突变

尽管利用上述实验得到了明确的结论，但有人认为在某些条件下，细菌的确存在类似定向突变的现象。凯恩斯（Cairns）等的研究结果显示，当以乳糖为唯一碳源时（在这样的条件下发酵乳糖的能力是有益的），大肠杆菌从不能发酵乳糖的 Lac^- 型转变成 Lac^+ 型的频率要比用葡萄糖作为碳源时高（发酵乳糖的能力是无益的）。

在这种情况下，突变似乎不是完全随机的。删除有益突变比无益突变更容易发生。但这并不意味着拉马克是完全正确的。这种效应仅出现于选择条件妨碍生长的情况下（即在细菌不被杀死的情况下，如噬菌体感染实验），并且与无法生长的细胞中发生的变化有关。细胞生长力的不足并不意味着像 DNA 复制（如 DNA 修复）等所有代谢活力的缺乏。有证据表明，细胞在这种状态下更易突变，这也能够解释所观察到的结果；但对该效应的解释还存在很大的争议。

2.2　突变的类型

2.2.1　点突变

有多种方法能够使遗传物质结构发生改变。利用在 DNA 序列中的某一位点发生改变的简单突变（点突变）已经建立了遗传学的许多基础理论。在点突变中，另一核苷酸代替了原核苷酸，即碱基替换。这种变化的结果取决于其本身的特性和位置。如果突变发生在基因编码区（该区域最终被翻译成蛋白质），将导致氨基酸序列改变，进而可能会影响蛋白质的功能。有些改变影响很小或没有影响，例如，改变的密码子如果依然编码相同的氨基酸，或新氨基酸与原氨基酸性质相似，则对蛋白质功能影响不大。

如三联体密码子 UUA 编码亮氨酸，单一碱基的改变会得到另外 9 个密码子中的一种，如图 2.4 所示。其中两个突变（UUG 和 CUA）是完全沉默的，依然编码亮氨酸。因此 UUA、UUG 和 CUA 这三个密码子是同义密码子。另两个突变（AUA 和 GUA）对蛋白质影响很小，因为它们编码的氨基酸（异亮氨酸和缬氨酸）与原本的氨基酸（亮氨酸）性质相似（均为疏水氨基酸）。苯丙氨酸（UUU 和 UUC）也是疏水氨基酸，但它在突变位点会引起蛋白质结构的显著改变。如果突变成 UCA，将由丝氨酸代替亮氨酸（两者性质显著不同），突变的结果由该氨基酸及邻近氨基酸在整个蛋白质的功能及构象中所扮演的角色而定。

最后两个突变（UAA 和 UGA）形成终止密码子（另一个终止密码子是 UAG），由于没有相对应的带有反义密码子的 tRNA，将导致翻译终止。这类突变（即无义突变）将产生被截短的蛋白质，基于被截短的程度不同，蛋白质可能具有或不具有功能。UAG 终止子也叫做"amber"（琥珀密码子），是德语"Bernstein"（琥珀）的英文翻译，而 Bernstein 是 UAG 的发现者之一。因此 UAA 和 UGA 终止子也分别被戏称为"ochre"（赭石密码子）和"opal"（乳白密码子），但这两个名字不常用。

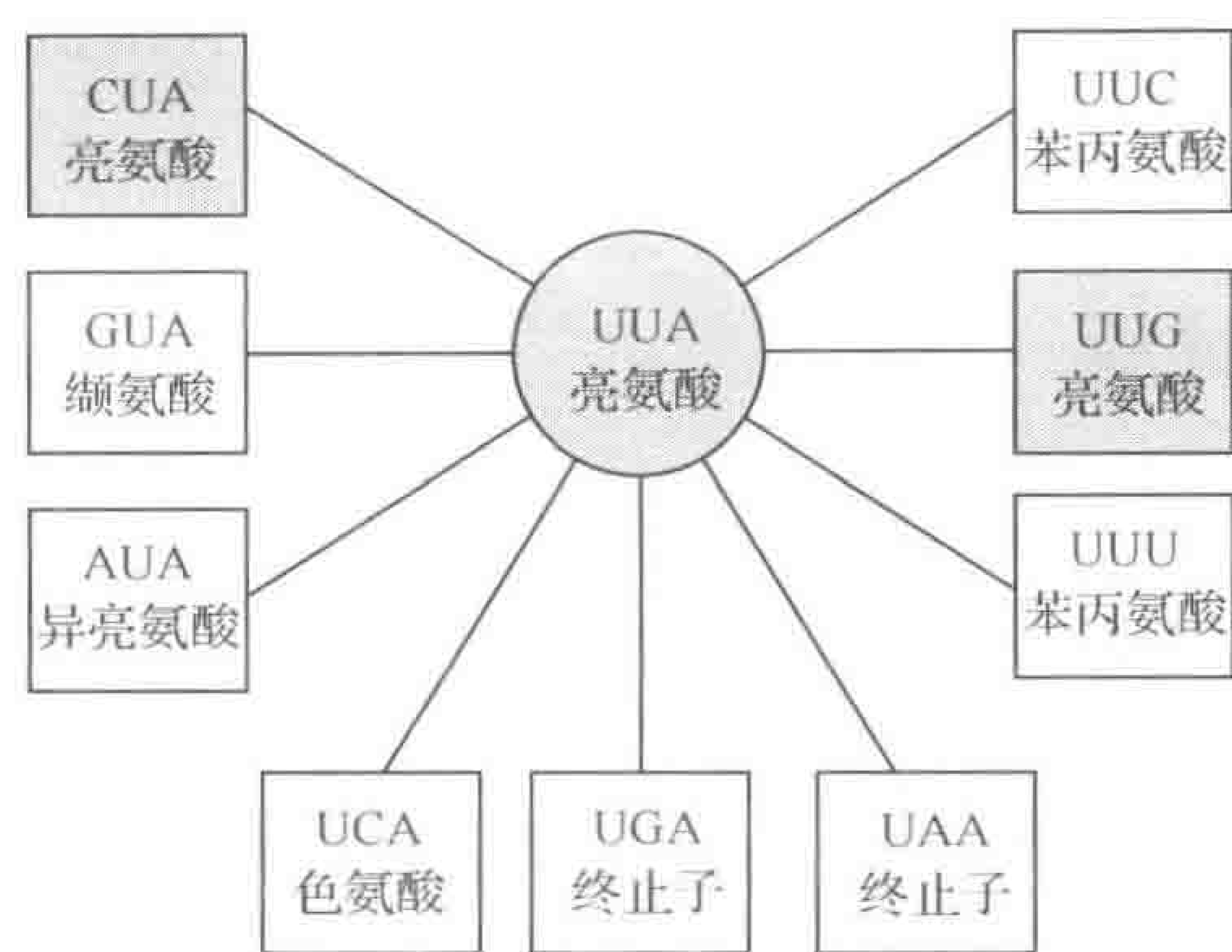


图 2.4 UUA 发生单碱基替换产生的密码子
沉默突变以灰色显示

另一个不同类型的突变也涉及单一位点改变，即在某一位点缺失或增加一个核苷酸(或多个核苷酸，但不是 3 的倍数)，基因的阅读框从该位点后发生改变，即移码突变。遗传信息的阅读框以三联体密码子为单位，中间没有标点符号间隔(阅读框只由翻译的起始密码子决定)，阅读框的改变将导致从突变位点开始合成完全不同的蛋白质，如图 2.5 所示。

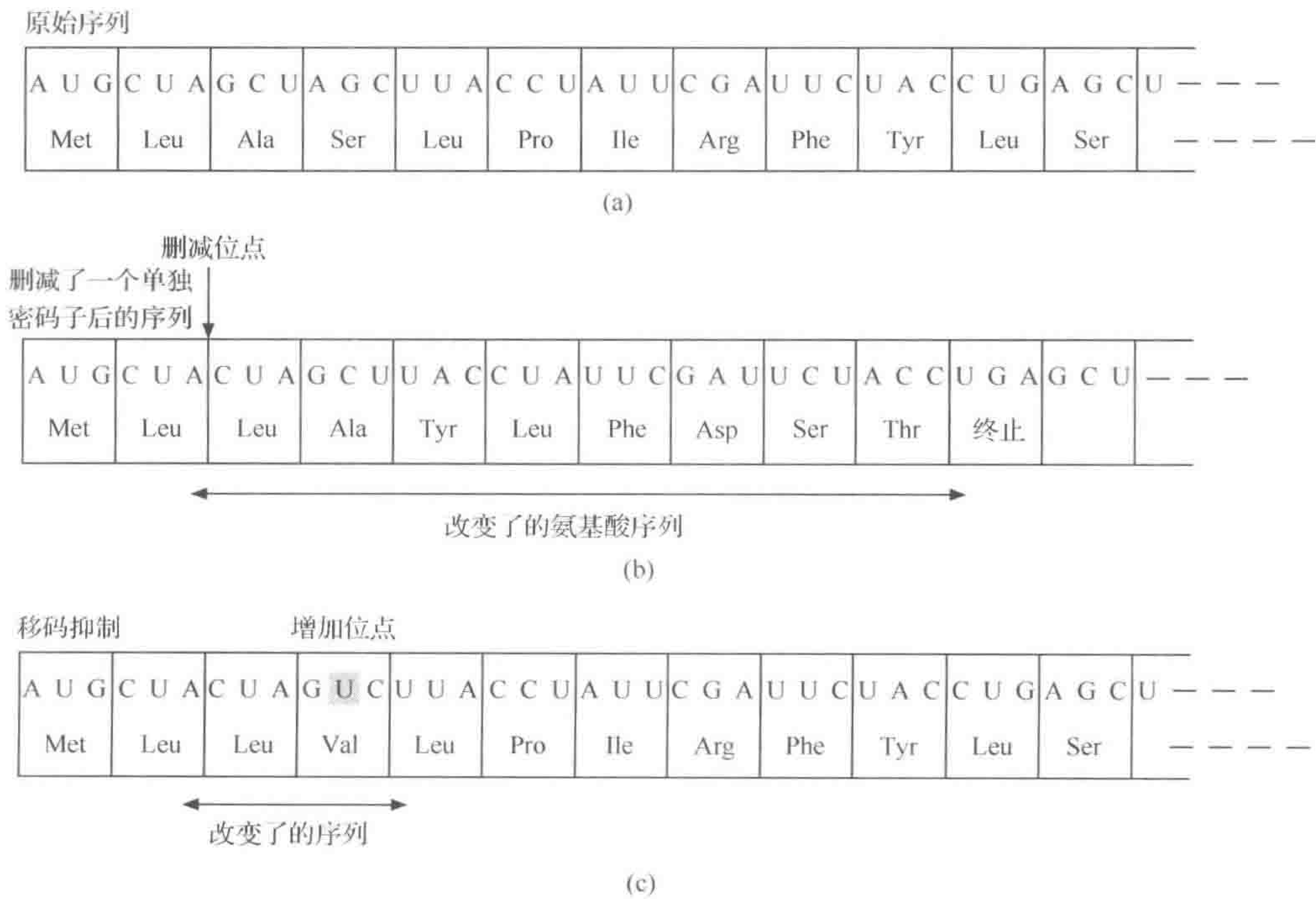


图 2.5 移码突变及抑制
(a) 原始序列(mRNA)及翻译产物；(b) 删除一个碱基改变了其后的阅读框，产生不同的氨基酸序列，并出现终止子；(c) 在不同位置加一碱基恢复了原始阅读框，可抑制突变

实际上,蛋白质的翻译在缺失位点之后不久即终止。对于多数基因,另两个可选的阅读框被终止密码子所终止,以防止错误翻译产生异常蛋白质。非翻译区也常含有终止密码子,而一个阅读框(如一个可读框或 ORF)是不含有终止密码子的较长 DNA,因此可利用其 DNA 序列的特点鉴定未知基因(参考第 10 章)。

如果点突变导致翻译过早终止,可能影响同一操纵子下游其他基因的表达。这种效应即极性突变,需进行遗传学分析。

2.2.2 条件性突变

虽然很多基因不影响细菌对抗生素或噬菌体的抗性,主要代谢产物的生物合成或碳源的利用。但其中有些基因是细菌必不可少的,任何与其相关的缺失突变体都将导致细菌的死亡(或无法生长)。涉及控制细胞必需功能(如 DNA 复制)的许多基因,通过遗传学方法分析基因及其产物的功能是十分重要,如可通过条件突变体进行分析。基因通常在某种条件下才具有正常功能,当条件发生改变,则功能缺失。相关功能的温度敏感性突变是极有应用价值的条件突变。例如,某一突变体,与 DNA 复制有关的基因发生温度敏感突变,使细菌在 30℃(允许温度)下能正常生长,但在较高温度下,如 42℃,则不能生长。本章后部列举了另一个条件突变体的例子,一种噬菌体突变(琥珀型突变),这种噬菌体只能在携带有琥珀突变抑制子的突变宿主菌内才能复制。

2.2.3 大片段 DNA 改变造成的变异

以上讨论的都是点突变,但是细菌(及其他有机体)发生的变异多数是由于 DNA 结构的大量改变造成的。其中最简单的是一个(甚至几个)基因的大部分缺失,从而使该基因功能完全丧失,由于这种缺失不可逆,因此对遗传分析十分有益。遗传分析通常是研究细胞发生的小概率事件(如两个紧密相连的基因进行重组),约 10^9 个细胞中发生一次,发生频率也依赖于使用的遗传标记的稳定性。点突变的回复率要高于此,原因却难以解释。

其他类型的大片段 DNA 改变对微生物的遗传多样性具有重要意义。当一段外源 DNA 插入某基因时,通常会使该基因失活。这种能够特异性地插入其他 DNA 序列,而产生插入突变,称为插入序列(insertion sequence, IS)。多数自发突变不是由于复制错误造成的,可能由于插入了一个插入序列而使该基因失活。

转座子与插入序列单元基本相似,由于二者都具有从一个位点移动(转座)到另一位点的能力,不同之处在于转座子携带有一个或多个可鉴定的遗传标记。研究最广泛的是带有抗生素抗性基因的转座子,其在进化及抗生素抗性传播中扮演重要角色。

有些 DNA 不是从一个位点移动到另一个位点,而是在原位进行倒位。如果倒位区域包含邻近基因表达所需信号,这种倒位将以一种容易逆转的方式(仍然遗传)打开或关闭相关基因。这种效应的最好的例子是鼠伤寒沙门氏菌(*Salmonella Typhimurium*)鞭毛抗原的变异。在第 7 章中,将对插入序列、转座子和倒位序列进行更全面的讨论。

通过对同一种的不同菌株全基因组序列分析发现,广泛存在的大片段 DNA 改变造成的变异正是菌株变异的来源,这一内容将在第 10 章继续讨论。在不同菌株间,不仅插

入和删除广泛存在(该位点常称为插入和缺失位点,因为很难断定在一个菌株中是插入,还是在另一个菌株中是删除),而且在不同菌株中,发现了不同的分布部位和(或)以不同方向定位的大片段基因。基因组的保守程度被称为同线性。

2.2.4 染色体外遗传因子及水平基因转移

除了染色体 DNA 结构的改变,细菌还常由于获得(或失去)染色体外 DNA(以质粒或噬菌体的形式)而产生变异。这些染色体外遗传因子控制许多特性,尤其像抗生素抗性。遗传信息从一个菌株转移至另一个菌株称为水平基因转移,在某些物种中,这种转移可能影响其染色体结构及染色体外遗传因子的获得。第 5 章将详细讨论质粒,第 6 章将讨论水平基因转移。

2.3 重组

长期以来,重组可在杂交后代对可见特征进行重分配,它一直是遗传学的基础性概念。“重组”一词也可用于细菌遗传学,即产生综合双亲特征的子代菌,也可指对 DNA 分子的断裂和连接。

举一个最简单的例子:分别使两个线性 DNA 分子从某一位点断裂,再交叉连接,产生两个重组 DNA 分子,这两个重组 DNA 分子分别含有两个亲代 DNA 的部分序列(图 2.6)。“重组”是一个广泛适用于重组机制的概念,其中最重要的是一般性或同源重组。同源重组要求重组的序列间具有较高同源性,因此任何两个同源 DNA 片段都可进行重组。反之,位点特异性重组对同源性要求不高或可无同源性,但如其概念所示,只能对特定序列进行重组。这两种重组方式的一个重要差异在于,同源重组需要蛋白质 RecA 参与,而位点特异性重组不需要。



图 2.6 两个线性 DNA 分子进行重组

同源重组是细菌间许多种基因转移(水平基因转移)的一种基本特征,即一个细菌将一部分染色体转移给另一个细菌。被转移的 DNA 片段不能独立复制,必须同受体染色体进行重组才能复制和遗传。值得注意的是,重组除了组合外源 DNA 外,在细胞中还有其他作用,尤其值得注意的是其参与 DNA 的修复(见第 1 章及本章其他部分),这对细胞来说非常重要。这也可能是细菌采取几种不同机制进行 DNA 分子重组的真正原因。

2.3.1 一般性(同源)重组过程的模型

假设一对含有同源区域的双链 DNA 分子相互配对,在一个 DNA 分子的一条链上产

生一个切口，另一个 DNA 分子的相对应区域将替代这条链。被替换的链随之断裂[图 2.7(b)]，形成部分交换的中间结构，断裂处重新连接形成交叉[图 2.7(c)]，即 Holliday 交叉(由 Robin Holliday 最初提出的模型)，另外两条链被切开并重新连接。为了更清楚地解析这种结构，其他的链必须被切断并重新连接。为了使之形象化，将 DNA 链进行旋转使之形成图 2.7(d)所显示的结构：因为 DNA 分子是螺旋状而不是线性的，所以该图与图 2.7(c)没有区别。将 DNA 链在箭头标记处切开，则产生来自双亲的一些遗传标

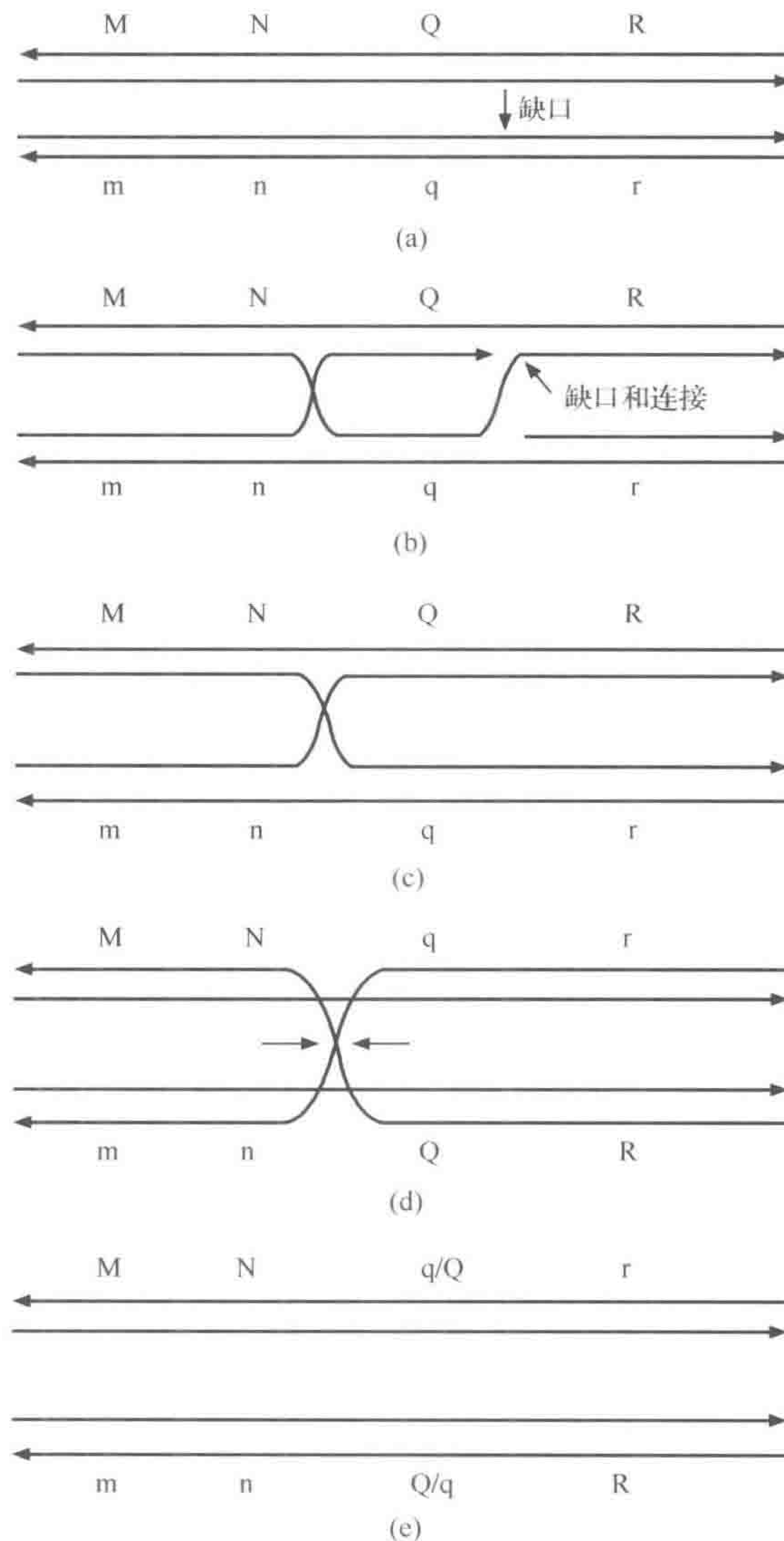


图 2.7 同源重组起始阶段

(a) 同源区域配对；(b) 产生切口链侵入另一 DNA 分子，替代了相对应的链；(c) 被替换的链产生切口，交换的链重新连接；图(c)和(d)一致，只是将右半部绕交叉点(箭头所指)旋转以说明下一阶段；(e)交叉链断裂并重新连接后，产生重组 DNA 分子

记的最终产物[图 2.7(e)],从而在遗传及分子水平发生重组。值得注意的是,图中所示的含有 Q/q 标记的短区域是异源双螺旋,即两条链分别来自于两个亲代。这种异源双螺旋要么被修复(即 q 转变成 Q,或相反),要么在修复前就被复制,产生综合了两个亲代特征的后代。

以上是忽略了 DNA 三维结构及可能产生不同产物影响的一个简单重组模型。重组过程包括 DNA 链的剪切,双螺旋解旋及再结合。

2.3.2 重组过程中的酶

重组过程的关键酶之一是 RecA 蛋白。在重组的背景下,RecA 蛋白的作用是促进重组 DNA 分子间的相互作用。RecA 蛋白在 DNA 链上多聚化,形成常见的螺旋状纤丝,使 DNA 螺旋呈伸展状态,从而促进其与其他 DNA 分子间的相互作用。由 *recB*、*recC* 和 *recD* 3 个基因编码的具有 3 个亚基的内切酶(RecBCD 内切酶)参与该过程。该酶是既有内切酶活性,又有外切酶活性的多功能酶,并能使 DNA 分子解旋,以提供必需的单链区域。当 DNA 分子解旋时,其中一条链被降解,直到 RecBCD 内切酶抵达一个特殊序列,即 *chi*(χ)位点。之后核酸酶降解活性被抑制,留下的一个单链 DNA 尾巴侵入其他链(在 RecA 的参与下)。因此,这些 χ 位点是重组热点。在大肠杆菌中,χ 位点序列为 5′-GCTGGTGG-3′(在其他细菌中可能有所不同)。如果随机分布,平均每 65kb(kilobase,千碱基)出现一次 8 碱基序列,而且通过染色体接合转移(参考第 6 章)产生的片段多数都足够长,可能包含一个 χ 位点。对于小片段 DNA,由于缺乏 χ 位点,会限制所观察到的重组的数量。例如,在基因转导中常出现该现象,在遗传操作实验(如基因替换)中则更为多见。

RuvA、RuvB 和 RuvC 3 个蛋白质在 Holliday 交叉处发挥作用。RuvA 结合在 Holliday 连接体处并稳定为后续事件所需的结构,RuvB(解旋酶)使附近 DNA 解旋,从而使 Holliday 交叉沿 DNA 移动(增加了异源双螺旋的长度)。RuvC(核酸酶)切割 DNA 链,以分解 Holliday 交叉结构。

如第 1 章所述,对 DNA 分子单链缺口进行修复时,需要另外一种不同的途径(尽管也依赖 RecA)。RecBCD 不参与此过程,作为替代,由 RecF 和其他几种蛋白质参与使 DNA 成为单链,并加载 RecA 以侵入姐妹染色单体。

2.4 表型

最初,人们通过其表型效应描述突变体,也就是其对细胞可见特征的影响。许多细菌突变株都可归为以下 3 种表型之一。第一类,需生长添加物的菌株,如添加某种特殊氨基酸。这是营养缺陷型菌株,其野生型、非突变菌株则为原养型。第二类,无法利用某种底物的突变株。最具有代表性的是无法发酵乳糖的大肠杆菌突变株(Lac$^-$)。第三类,抗生素抗性突变株。例如,编码一种核糖体蛋白的基因发生突变,使细胞对链霉素产生抗性。染色体基因突变和质粒的存在都可导致菌株对抗生素产生抗性,正确区分二者非常重要。

细菌对其他化学成分或噬菌体的抗性有利于对其进行遗传分析。突变也可以导致表型的其他改变，在第 9 章将对其进行介绍。

非常遗憾，对表型的观察并不是总能告知有关其基因型的更多信息，也就是说，无法得知究竟哪个基因被影响及该基因突变的性质。某一特定表型特征常受到许多基因的影响。例如，合成色氨酸的能力是由色氨酸操纵子(*trp* 操纵子)内 5 个相连的基因控制，其中任何一个基因缺失都将导致细菌无法合成色氨酸，那么该菌株就是 Trp^- 型。这并没有告知哪个基因有缺陷，但通过进一步分析可能表明具体是哪一个基因发生了变化。准确地描述菌株的特征对研究者来说非常重要。因此人们建立了描述细菌菌株表型及基因型的标准系统命名法(框 2.2)。必要时，每个基因可用一个附加的字母表示。例如，*trpA* 和 *trpB* 基因分别编码色氨酸合成酶的构成亚基，参与色氨酸合成过程中将吲哚-3-磷酸甘油转变成色氨酸的最后阶段。

框 2.2 遗传学系统命名法

在任何情况下，都用 3 个字母的缩写对细菌的基因及其产物按标准系统法进行命名。本书所涉及的基因参考附录 E。

表现型用普通字母表示(非斜体)，第一个字母大写，通常加一后缀说明变化的类型。如：

Amp^r 氨苄青霉素抗性菌株；

His^- 组氨酸营养缺陷型菌株；

Lac^- 无法发酵乳糖的菌株。

值得注意的是，第二和第三种突变类型容易混淆。Lac^- 菌株是不能利用乳糖，而不是不能合成乳糖。作为常识，应注意辨别这些可能的情况。

基因型用 3 个斜体字母表示，第一个字母不大写，通常在后面加一大写字母代表某一特定基因。例如，不能发酵乳糖的菌株是由于编码 β-半乳糖苷酶的基因有缺陷，其基因型记为 *lacZ*。描述基因型时不必加“—”号，因为突变本身已包含了这一意义，但为了阐述的更清楚，有时也用“—”或“+”。

该规则也适用于描述特定基因及其产物。例如，当提到“*recA* 基因”，那么该基因编码的蛋白质就称为“RecA 蛋白”。

如果一个基因被插入(尤指插入移动元件，如转座子或插入序列)而失活，以 *hisA* 基因为例，可表示为 *hisA*::Tn*10*，指 *hisA* 基因由于插入转座子 Tn*10* 而发生突变。这种表达方式也可表示菌株携带有能形成某种表现型的质粒。如果菌株的基因型是 *hisA*/F′*hisA*$^+$，则说明染色体 *hisA* 基因发生突变，但含有 F′ 质粒，该质粒携带有 *hisA*$^+$ 基因(用“+”号代表)。

该规则并不是广泛适用的。本书中很多噬菌体基因不用三字母表示，在真核生物中基因的命名更是多种多样。但用斜体字母表示基因，用非斜体字母表示基因产物是基本原则。

如果将某一菌株命名为 *trpA*，并不意味着它一定是营养缺陷型，只是代表 *trpA* 基因发生改变而已。在不使基因编码产物失活的情况下，基因发生改变的方式有许多种，如突变赋予了相关基因产物对温度的敏感性。

还有些基因被记为 *trp*，但它们并不是 *trp* 操纵子的组成部分，这些基因突变不会产生营养缺陷型。例如，*trpT* 基因编码色氨酸特异的 tRNA，*trpS* 基因负责编码色氨酰 tRNA 合成酶，而 TrpR 蛋白是一种调控 *trp* 操纵子的抑制蛋白。*trpR* 突变可能增加色氨酸操纵子的表达，或者当色氨酸存在时，降低细胞关闭 *trp* 操纵子表达的能力。

随着基因组测序时代的来临，许多基因仅通过序列就可被鉴定，但很难确定其功能。一般而言，可通过同已知序列对比来推测其功能，但还有大量尚未知其功能的假基因。在第 10 章中将继续讨论这个问题。

2.4.1 表型修复

回复和抑制

点突变是由 DNA 单一位点随机改变所造成的，因此有可能发生第二次突变，从而恢复到原始 DNA 的序列（或是一条实际效果不明显的 DNA 序列）。如图 2.4 所示，带有 UUU 密码子（编码苯丙氨酸）的突变株可能进一步突变，回复成 UUA 密码子（编码亮氨酸，真正的回复突变），或发生一个碱基替换变成 UUG 或 CUU（均编码亮氨酸）。突变后的菌株同原始菌株特征相同，即所谓的回复突变。

突变对细菌的影响可被第二个不相关的突变消除，即抑制。抑制可以以多种方式发生，多数方式对于所涉及的基因是特异的，即当突变的第二个基因能对抗第一个基因功能缺失的不利影响时才出现抑制。以下两种抑制类型较为重要。

第一，移码突变形成抑制。可通过恢复删除的碱基来回复突变，也可通过增加或删除其他碱基（不必在原始突变的同一位点）实现抑制，但增加或删除的碱基应是 3 的倍数（见图 2.5）。从而使阅读框回复到原始阅读框，仅有少量密码子发生改变。这种产物的改变是否会对生物学功能产生影响取决于原始突变的抑制，当然这种抑制又是依赖于序列改变长度，特征和其对蛋白质功能的影响。

第二，伴随着无义突变，出现了另一种重要的抑制类型，即在编码序列中出现终止子。在正常情况下，由于没有与终止子对应的 tRNA，而导致翻译终止。tRNA 分子本身由基因编码，也易发生突变，使突变的 tRNA 容易（或也能）识别终止子，而不是它应该识别的密码子。如图 2.8 所示，原始突变将谷氨酸密码子（CAG）变成终止子（UAG）。要抑制该突变可通过改变编码谷氨酸 tRNA 的基因，使其编码的 tRNA 的反义密码子可识别终止子。谷氨酸重新插入到编码的肽链上，从而使终产物与野生型蛋白质相同。由于有多个编码谷氨酸 tRNA 的基因，细胞不会丧失识别正确的谷氨酸密码子的能力。

有人认为，这种抑制型 tRNA 会阻止翻译区内正常肽链翻译的终止而损害细胞。实际上，抑制型 tRNA 数量很少，核糖体只在极少数情况下能读通终止子。另外，在正常的终止位点，释放因子控制翻译的终止，并有多个终止密码子。综合以上因素，即使存在抑制型 tRNA，在真正的翻译终止位点翻译也会被有效终止。

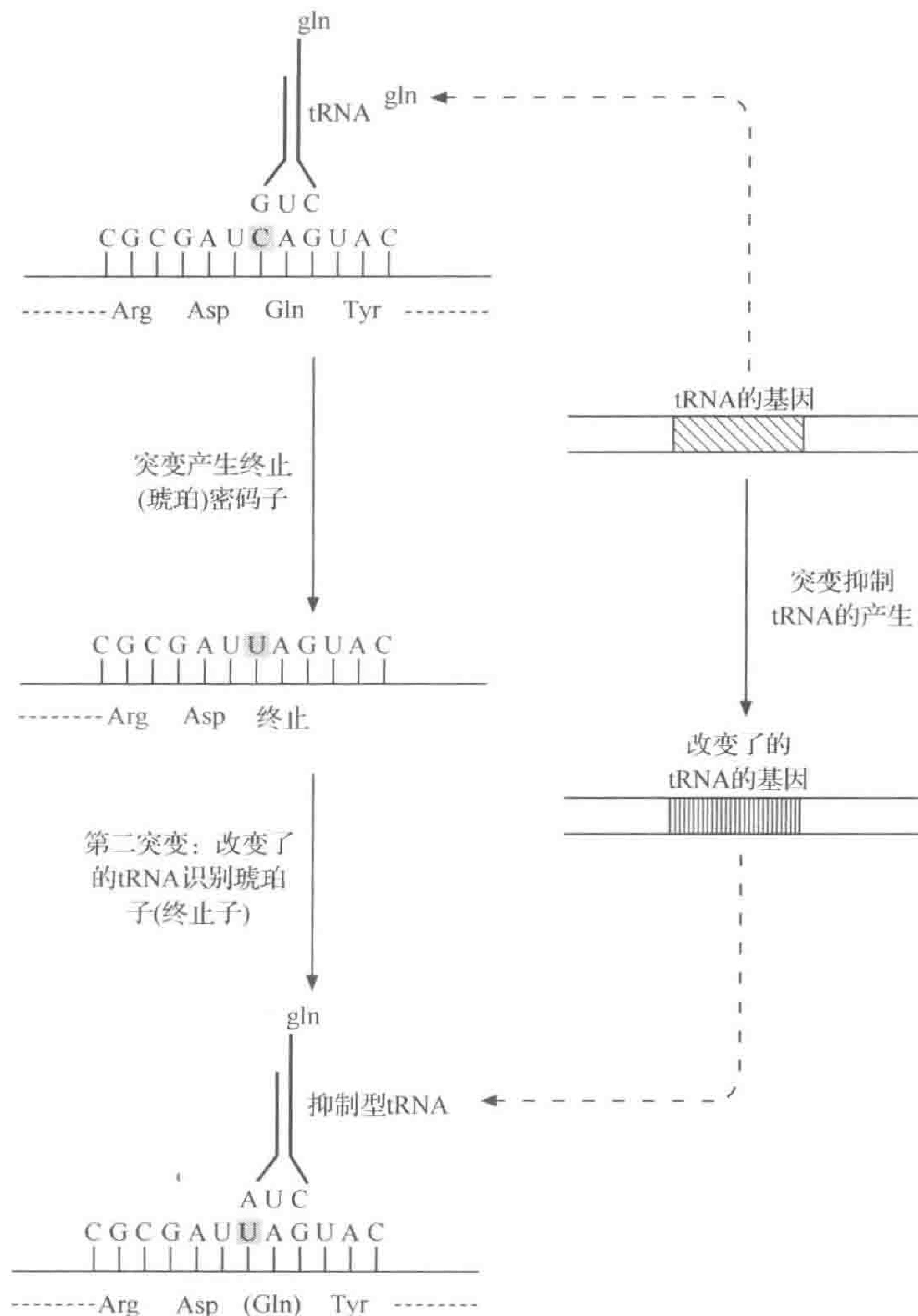


图 2.8 抑制无义突变

CAG 发生单碱基替换变成终止子 UAG，导致翻译提前终止；tRNA 的基因可突变产生能识别 UAG 的 tRNA 突变体，以抑制此无义突变

这种抑制类型的重要意义在于 tRNA 突变不仅能抑制原始突变，还能抑制任何相关的突变。当然，许多密码子通过突变都可形成不同的终止子(从而有一定数量的具有不同氨基酸特异性的抑制型 tRNA)，因此在多数情况下，抑制型 tRNA 将导致肽链中插入不正确的氨基酸。尽管可获得全长产物，但它可能不具备全部功能。除非可克服其缺陷，这在噬菌体遗传学中非常有用，如广泛应用的琥珀型突变(amber mutation)。带有琥珀型突变的噬菌体在正常的宿主细菌中显示突变表型(或不能生长)，但在具有琥珀型突变抑制的宿主细菌中则表型为野生型。后一种宿主细菌作为琥珀型突变噬菌体可使用的宿主。如本章前面所述，这是利用条件突变体的另一个例子。

互补

突变表型回复为野生型的另一条途径是获得携带有正常功能基因的质粒。例如，大肠杆菌 Lac⁻菌株缺乏功能性的 β-半乳糖苷酶，当引入一个携带 β-半乳糖苷酶基因(*lacZ*)的质粒后，则可以利用乳糖，即质粒可以与染色体缺陷互补。这只有在有功能的突变基因存在下才有效，即突变体是隐性的。

传统的遗传学研究路线为：首先，基于改变的表型分离突变体，再试图鉴定遗传改变的原因。基因克隆技术，尤其是基因组测序技术的出现，给人们提供了另一条研究路线。通过这些技术，可以获得某个有机体的多数甚至全部基因序列，但不知其功能。如果推断某一特定基因具有某种功能(如与其他来源的已知基因进行序列比较)，可以通过检测该基因补充特异突变体的能力来验证假设。

2.5 突变的机制

2.5.1 自发突变

自发突变产生于 DNA 复制过程中出现的错误。然而 DNA 的结构及其复制(见第 1 章)都不允许出现错误，那么自发突变是如何产生的呢？在正常的 DNA 分子双螺旋结构中，只允许碱基 A 和 T、G 和 C 配对，任何其他配对方式都导致双螺旋扭曲，这样的扭曲被酶移除并修复(见第 1 章)。

这是基本原则。如果碱基总以人们所设想的结构存在，则配对原则是正确的。但碱基还存在其他结构，即每个碱基都存在含有不同数量氢键的互变异构体。如图 2.9 所示的两组互变异构体，氨基-亚氨基异构体及酮-烯醇异构体(还存在其他类型的同分异构体，在此不做讨论)。碱基配对结果如图 2.10 所示。在正常情况下，氨基型腺嘌呤与酮型胸腺嘧啶配对[图 2.10(a)]。而亚氨基型腺嘌呤则与胞嘧啶配对[图 2.10(b)]，烯醇型胸腺嘧啶则与鸟嘌呤形成氢键[图 2.10(c)]，这些非正常配对形式处于热不稳定状态。如果考虑正常和非正常配对间的平衡，则在 $10^4 \sim 10^5$ 个分子中约有 1 个分子在某一位点出

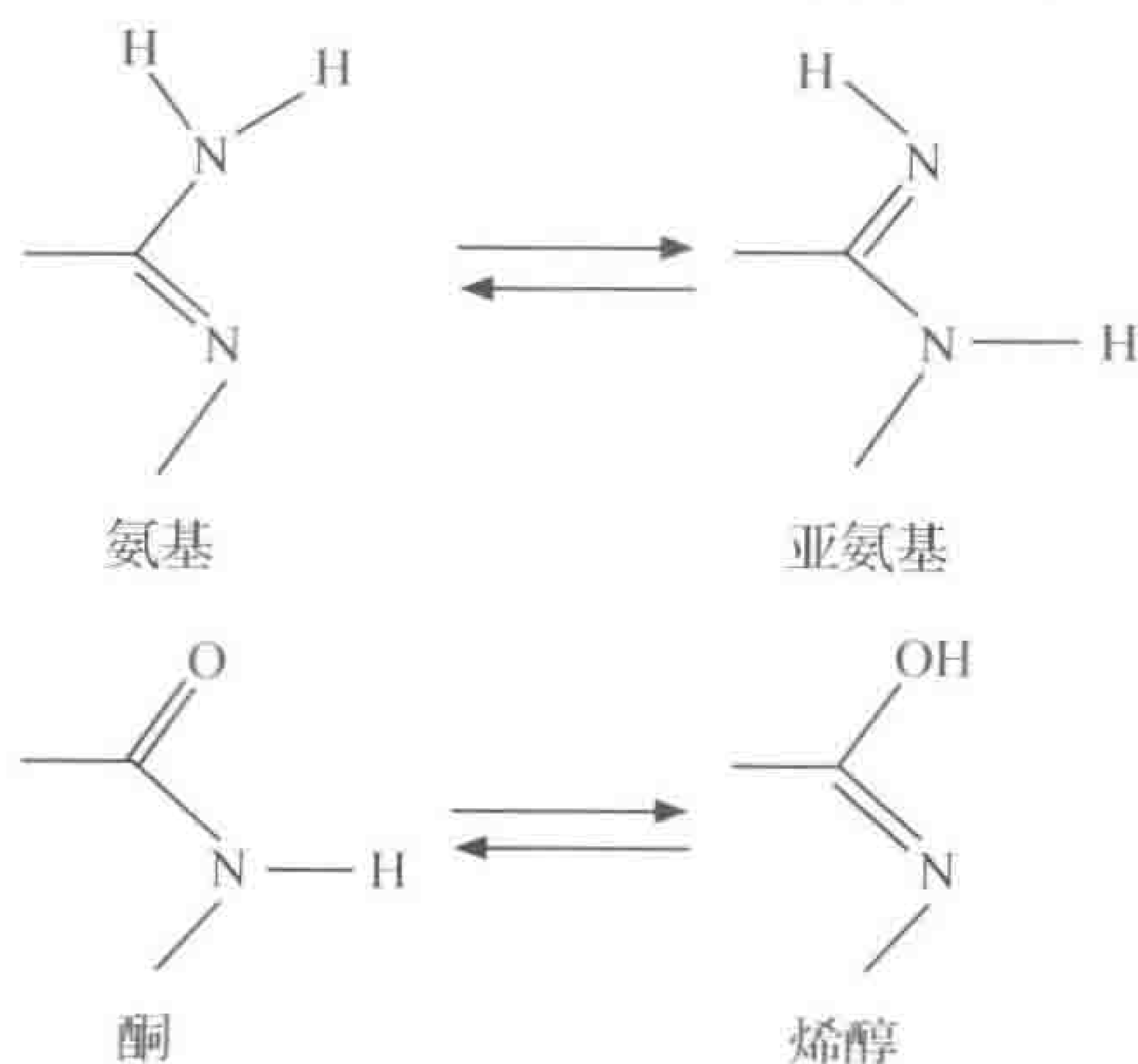


图 2.9 氨基-亚氨基及酮-烯醇同分异构体

现非正常配对。如果发生在 DNA 分子中就会导致该位点产生突变，每一代每一碱基其发生突变的频率为 10^{-5}～10^{-4}。

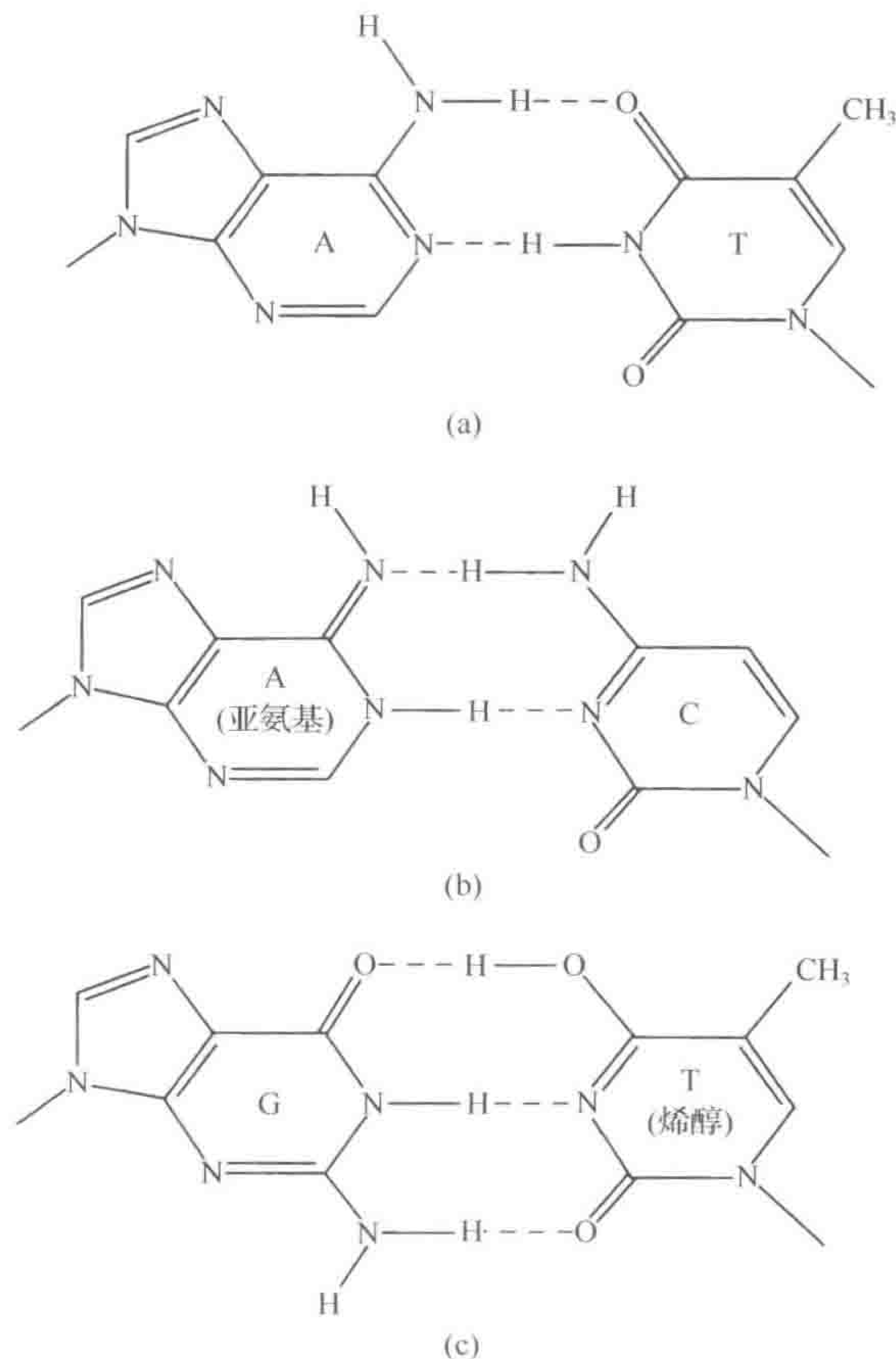


图 2.10 碱基的同分异构体导致错配

(a) 腺嘌呤(氨型)与胸腺嘧啶(酮型)的正常配对；(b) 亚氨基型腺嘌呤与胞嘧啶配对；(c) 烯醇型胸腺嘧啶与鸟嘌呤配对

这一问题的解释在于，实际观察到的突变频率远低于上述值(至少在大肠杆菌中如此)。原因是在 DNA 复制过程中，DNA 聚合酶的 3′→5′外切酶活性具有校正功能，可移除增长链的 3′端错配的碱基，如第 1 章所述。非正常的异构体被认为能迅速回复到能量更稳定、不能与原始链的相应碱基正确配对的状态。校正机制可移除错配，当校正失败时才产生自发突变。

事实上，校正机制失败并不足以形成突变，因为在对损伤或改变的 DNA 进行修复的几个连续发生的机制中，这只是第一种(见第 1 章及本章后面部分)。DNA 修复机制是细胞防御造成 DNA 损伤的各种因素的重要组成部分。校正机制及其他修复机制的存在保持了低突变率的发生。值得注意的是，不同的菌株之间，突变率各不相同。在某些细菌中，如增变菌株，突变率比普通菌株高几千倍。这些菌株经常含有 DNA 聚合酶的校正活性的突变(参见第 1 章)。另外，有些含有短的重复序列的基因易发生突变(复制过程中碱

基插入或缺失)，如熟知的同聚体使基因表达或关闭(参见第 7 章)。

2.5.2 化学诱变剂

由于自发突变率过低，难以分离各种类型的突变体(除了较易选择的突变，如抗生素抗性)，因此必须寻找增加突变率的方法。如果欲改变已知基因，可利用体外诱变剂或人工合成目标基因的方法(参见第 8 章)。有些情况下，使用化学或物理方法更适宜(或必须使用)，例如，要研究哪些基因与某一表型相关，就需要更为广泛的突变方法。转座子突变体是达到这一目标的另一种选择(参见第 9 章)。

许多不同的化学试剂能与 DNA 相互作用，或影响 DNA 复制，使 DNA 序列发生改变。这些试剂可对 DNA 碱基进行化学修饰，形成不同的碱基。例如，亚硝酸能引起氧化脱氨基作用，使氨基转变成酮，如胞嘧啶会转变成尿嘧啶(图 2.11)。尿嘧啶不是 DNA 分子中的正常碱基，细胞中的酶会将其移除。然而，如果在复制过程中它得以保留，则与腺嘌呤配对，使 C-G 配对变为 U-A 配对，最终变成 T-A 配对。腺嘌呤脱氨基产生次黄嘌呤，与胞嘧啶配对。

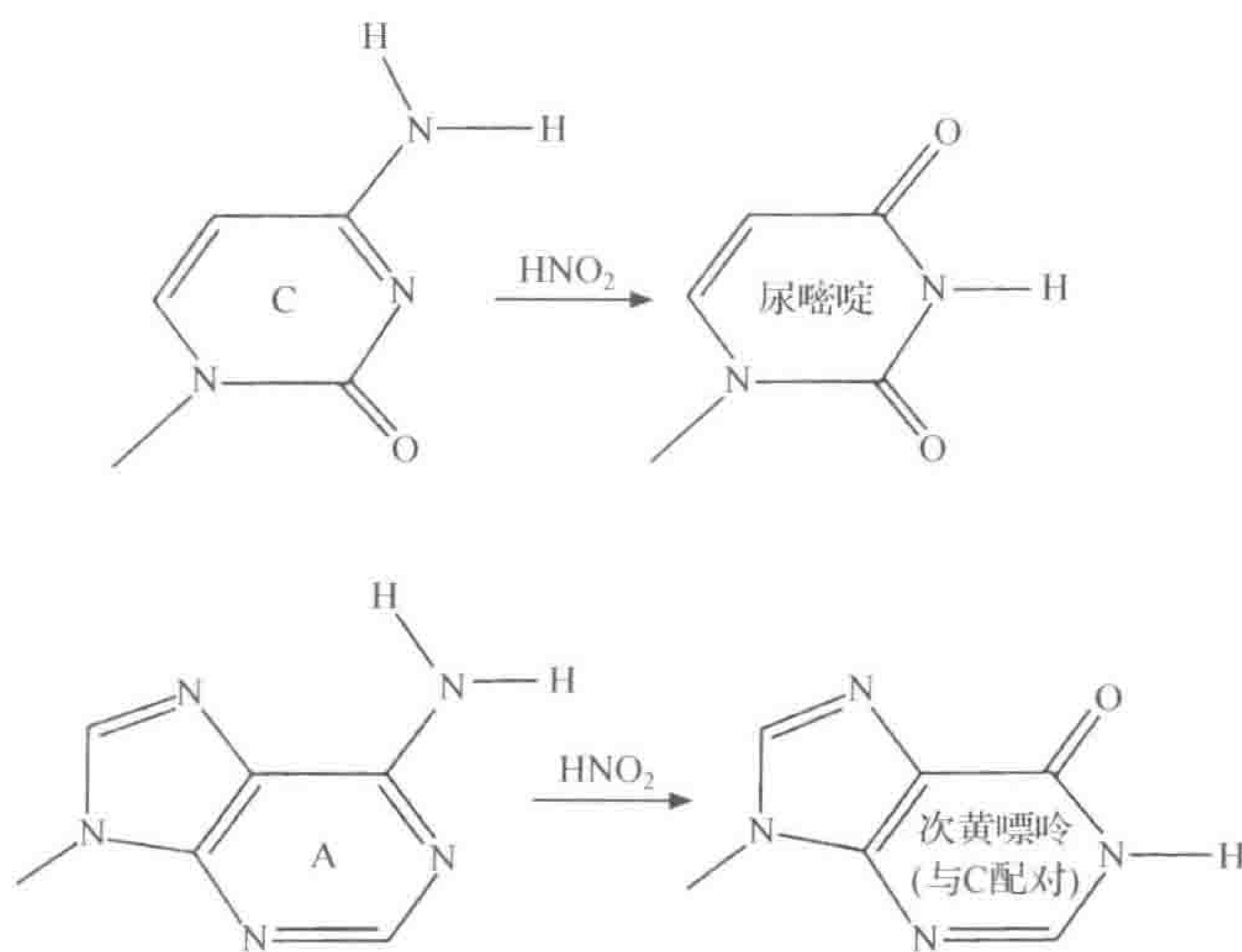


图 2.11 亚硝酸引起碱基的氧化脱氨基作用

亚硝酸可直接作用于分离的 DNA，将此 DNA 重新导入细菌细胞，则产生突变。在噬菌体及质粒中容易对 DNA 进行分离及再导入，因此这种处理方法十分有效，不必将整个细菌暴露于诱变剂就可获得突变的噬菌体或质粒。这保证了所选择的突变是由噬菌体或质粒 DNA 改变所造成，而不是由染色体基因改变造成的。

有些化学诱变剂不能作用于分离的 DNA，必须作用于细胞内 DNA。烷化剂是极强的诱变剂，如甲基磺酸乙酯(ethyl methane sulphonate, EMS)和 1-甲基-3-硝基-1-亚硝基胍(1-methyl-3-nitro-1-nitroso-guanidine, MNNG)，因此使用时也更危险，尤其是后者。它们可使核苷酸在任意位置发生烷基化，尤其是鸟嘌呤的第 6 位氧原子，从而在复制叉附近引起多重密集的连锁突变。

一种类型为插入型试剂，如吖啶橙和溴化乙啶作用方式不同。这些分子含有一个扁平的环状结构(图 2.12)，可插入到相邻碱基的双螺旋核心中，使 DNA 复制时增加(有时缺失)一个碱基，而产生移码突变。这些染料(尤指溴化乙啶)与 DNA 形成的复合物显荧光，因此也常用于分子生物学实验中检测 DNA 分子。

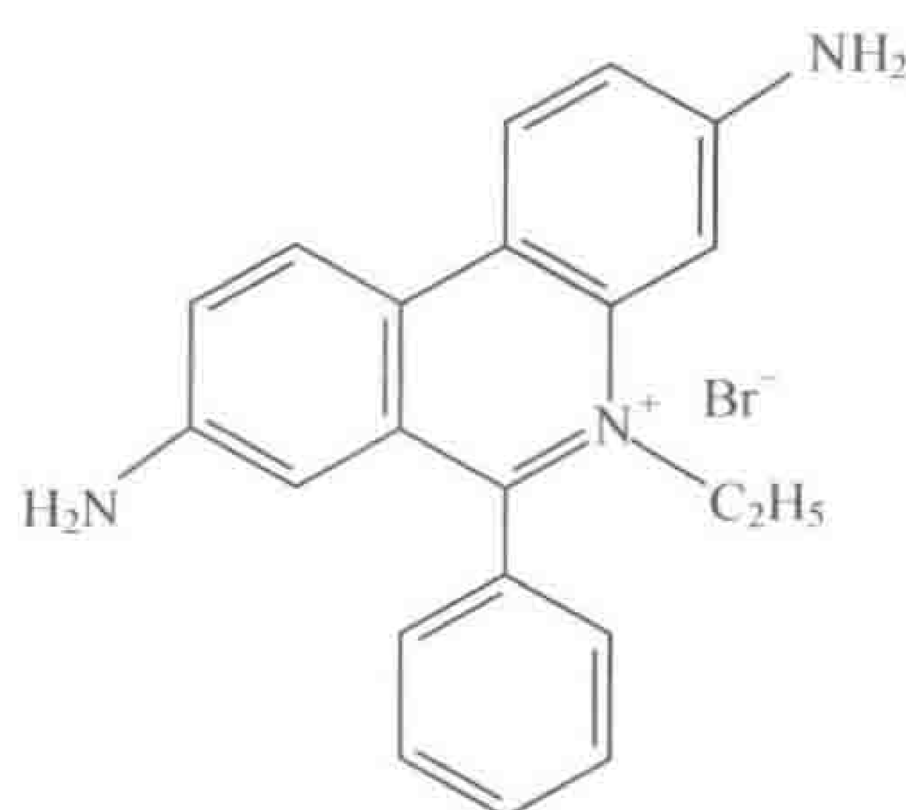


图 2.12 溴化乙啶

另一类型的试剂为碱基类似物，只作用于生长的细胞，例如，5-溴尿嘧啶是胸腺嘧啶的甲基被溴原子取代而形成的胸腺嘧啶类似物(图 2.13)。5-溴尿嘧啶能同模板链上的腺嘌呤形成碱基配对，从而替代了 DNA 分子中的胸腺嘧啶。由于上述同分异构现象在 5-溴尿嘧啶中很常见，因此在复制过程中，相对于腺嘌呤它更易与鸟嘌呤配对，从而产生 A-T 被 G-C 取代的突变。

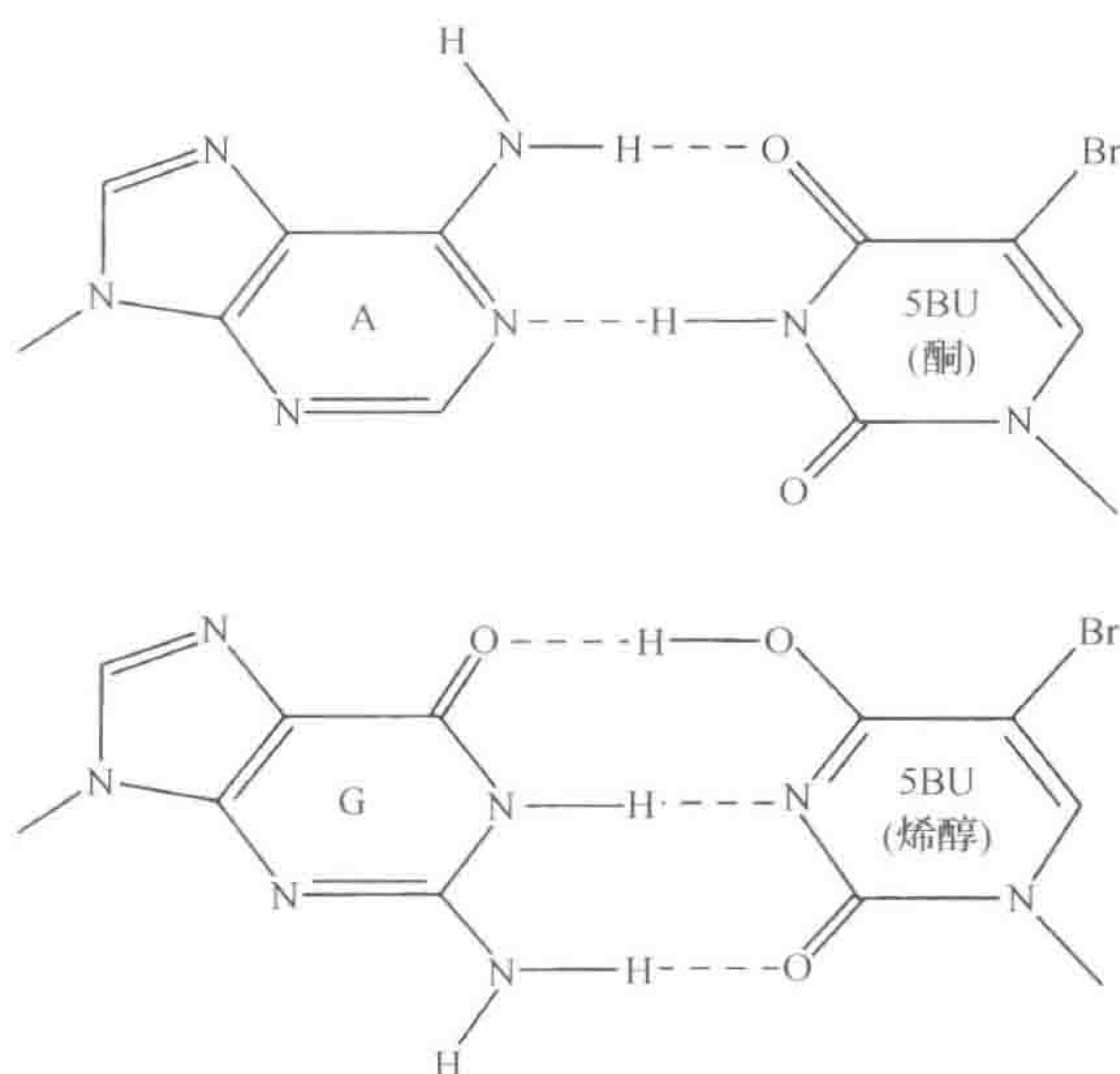


图 2.13 碱基类似物 5-溴尿嘧啶增加同分异构体形式

2.5.3 紫外线照射

理论上任何损伤 DNA 的化学试剂要么导致生物体的死亡，要么突变者得以幸存。辐射也造成同样的结果。目前已有多种辐射方法用于产生突变。高能量射线，如 X 射线

和γ射线需要昂贵的仪器及安全设备，尽管它们在过去得到应用，但对进行常规实验的微生物学实验室并不适用，而且它们还引起微生物不易修复的染色体损伤。另外，紫外线(UV)较易控制(对眼睛和皮肤的保护是必要的)，需要的仪器也相对便宜。

人们所关注的紫外线照射的主要效应是能产生嘧啶二聚体(通常指胸腺嘧啶二聚体，尽管胞嘧啶也常有二聚体出现)。在紫外线照射下，一条DNA单链上两个相邻的嘧啶间共价连接(图2.14)，形成的嘧啶二聚体不能被复制，而导致细胞死亡，除非细胞能修复该损伤。细胞试图修复由紫外线照射引起的损伤可产生类似诱变剂的效应。请参考紫外线照射DNA损伤修复机制。

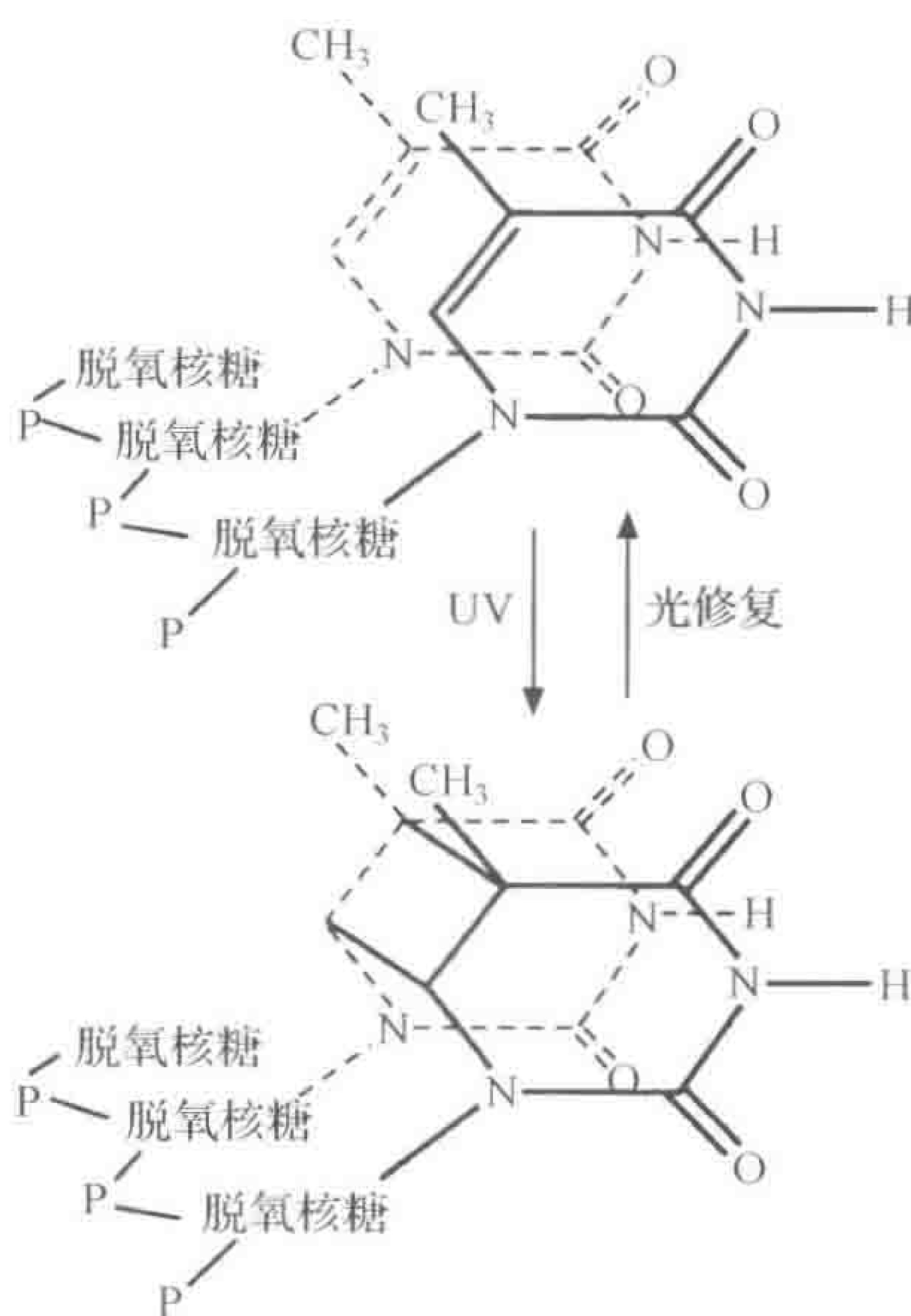


图2.14 胸腺嘧啶二聚体结构

光修复

光修复是抵御紫外线照射损伤的最佳办法。在可见光下，经酶催化(光裂合酶)打开两个嘧啶间的共价连接，使该位点处重新恢复了原始碱基序列。光修复方法十分有效，不导致突变。因此，在实验室利用紫外线照射进行突变时，细胞经紫外线照射后，在恢复期应将其避光(如可将培养瓶用锡纸包裹)。

SOS修复

避光后阻止了光修复，但细胞转而利用其他纠错修复过程，如剪切修复及重组修复(复制后修复)，参见第1章。只有当这些修复过程被另一种"错误倾向"的修复机制所压制(或使用缺乏纠错修复的突变株)，才能得到显著数量的突变株，即SOS反应。(尽管人们认为SOS反应与紫外线照射有关，但它也参与其他类型的DNA损伤修复，并与压力反应系统有关。)

当DNA损伤时，大量参与DNA损伤修复的基因被诱导表达。包括剪切修复基因*uvrA*、*uvrB*和*recA*，以及参与"错误倾向"修复的基因。可以通过辐射λ噬菌体，检测其

在被辐射及未被辐射的宿主细胞中形成噬菌斑的能力来展示损伤修复的这种可诱导性。在感染噬菌体之前用低剂量紫外线照射的宿主细胞形成更多的噬菌斑，这是由于紫外线照射引起的 DNA 损伤，在宿主细胞中诱导表达了参与损伤修复的酶。

这种诱导机制诱导表达了两种基因的产物，RecA 和 LexA。LexA 蛋白是 SOS 反应相关基因的阻遏蛋白，这些基因当然也包括 *recA* 和 *lexA* 本身。当 RecA 与 LexA 蛋白结合时，LexA 蛋白具有自身水解功能。RecA 蛋白以协同酶的作用促进 LexA 自身的水解。在该过程中，RecA 蛋白与损伤形成的 DNA 单链结合后，RecA 发挥其活性。蛋白质构象的改变使其与 LexA 结合，导致 LexA 的裂解及 SOS 基因的表达。

另外两个 SOS 基因，*umuC* 和 *umuD* 也参与突变形成，缺乏这些基因的菌株不仅对紫外线更敏感，还不易发生紫外线诱导的突变。某些质粒携带相似基因（*mucA* 和 *mucB*），这些基因的存在增强了细菌对紫外线的抗性，也导致紫外线及其他许多诱变剂诱导的突变水平增加。当存在 DNA 损伤时，DNA 聚合酶（DNA 聚合酶Ⅲ）失活，UmuCD 复合物可替代聚合酶行使其功能。UmuCD 聚合酶的低特异性使其能越过损伤处继续进行 DNA 合成，从而在新生成的 DNA 链中产生错误。

2.6 突变体的分离与鉴定

2.6.1 突变与筛选

根据以上方法，可方便地对 DNA 进行随机改变。但随之而来的问题是如何分离特定基因改变的突变体。如果可能的话，最有效的分离方法是在一定生长条件下野生株或突变株之一不能生长。应依据突变的类型及对细胞产生的影响筛选突变株。参照以前内容，可将突变分为下述 3 个类型。

1. 能抵抗抗生素或特异噬菌体、有毒化学物质及其他任何对亲本细胞有致死或抑制作用的突变体。

2. 营养缺陷型，也就是突变体生长需要添加其他生长因子，如氨基酸。

3. 不能利用某一特殊生长底物（通常为糖）的突变体。

人们能获得任何细菌的抗性突变体（只要它能在实验室条件下生长）。但第二和第三种类型的突变体则不易分离，除非细菌能在简单的限制培养基中生长。这也是为什么细菌遗传学家最初只对大肠杆菌和其他有限范围的细菌，如鼠伤寒沙门氏菌和枯草芽孢杆菌（*Bacillus subtilis*）进行研究的主要原因，尽管如今科学家所研究的细菌范围已大大增加。

理论上，抗生素抗性的突变株很容易分离。简单地说，将培养液涂布于含有一定浓度抗生素的琼脂平板上，亲本细菌的生长则受到抑制。这种选择性条件十分有效，不必使用任何诱变剂。由于使用的细菌数量极大（10^8～10^9 个，或 0.1～1mL 过夜培养的大肠杆菌），即使是低频率的自发突变也能被轻易地检测出来。如果突变频率为 10^{-8}，将 1mL 含有 10^9 个细菌的培养液涂布到平板上，每个平板上则平均出现 10 株抗性突变株。

由于很难找到突变体能生长，而亲本细菌不能生长的条件，因此通常不能通过这种方法将营养缺陷型的突变体直接筛选出来。例如，野生型大肠杆菌能在含有无机盐、糖（作

为碳源或能量)、铵离子(作为氮源)的基本培养基中生长。组氨酸营养缺陷型突变株缺失了一种组氨酸生物合成途径中所需的酶,因此需要在该培养基中额外添加组氨酸才能生长。而亲本细菌在含有组氨酸的培养基中也会快速生长。因此,尽管不可能设计出只让营养缺陷型突变株生长的培养基,但让营养缺陷型突变株不能生长却是容易的。

如果突变率足够高的话,可以逐一挑取单克隆菌落,鉴定其在未添加任何物质的基本培养基上能否生长。在细菌遗传学实验室中常使用这种方法用于多种用途。用灭菌牙签挑取单菌落,将其依次接种在不同培养基表面,以检测其在不同培养基上生长的能力。将菌落系统地接种在有序号的带格子的平板上,不同平板之间就可以相互比较,这样每一平板上可鉴定多达 100 个菌落。如果用这种方法对几百个菌落进行鉴定,劳动强度很大(尽管现在可应用自动控制程序,包括机器人工作站)。即使经诱变剂处理,获得理想突变体的频率仍然很低,因此通常要检测成千上万,甚至几百万个菌落才能获得。

2.6.2 影印平板法

本章前面部分涉及影印平板技术,用包有丝绒衬垫可将原始平板上所有菌落复制到几个平板上,这可以在大量菌落中快速地检测突变株,如营养缺陷型突变株(图 2.15)。在实验中,将经遗传突变操作的培养液涂布到营养培养基平板上以获得单克隆菌落,在此平板上突变体和亲代细菌均生长。经培养后,用灭菌包有丝绒衬垫将所有菌落影印在基本培养基平板上,以及含有添加物的另一平板上(以筛选组氨酸营养缺陷型突变株为例,添加物为组氨酸)。需要组氨酸的营养缺陷型突变株在第一个平板上不能生长,而在第二个平板上生长,由此可将其筛选出来进行进一步研究。

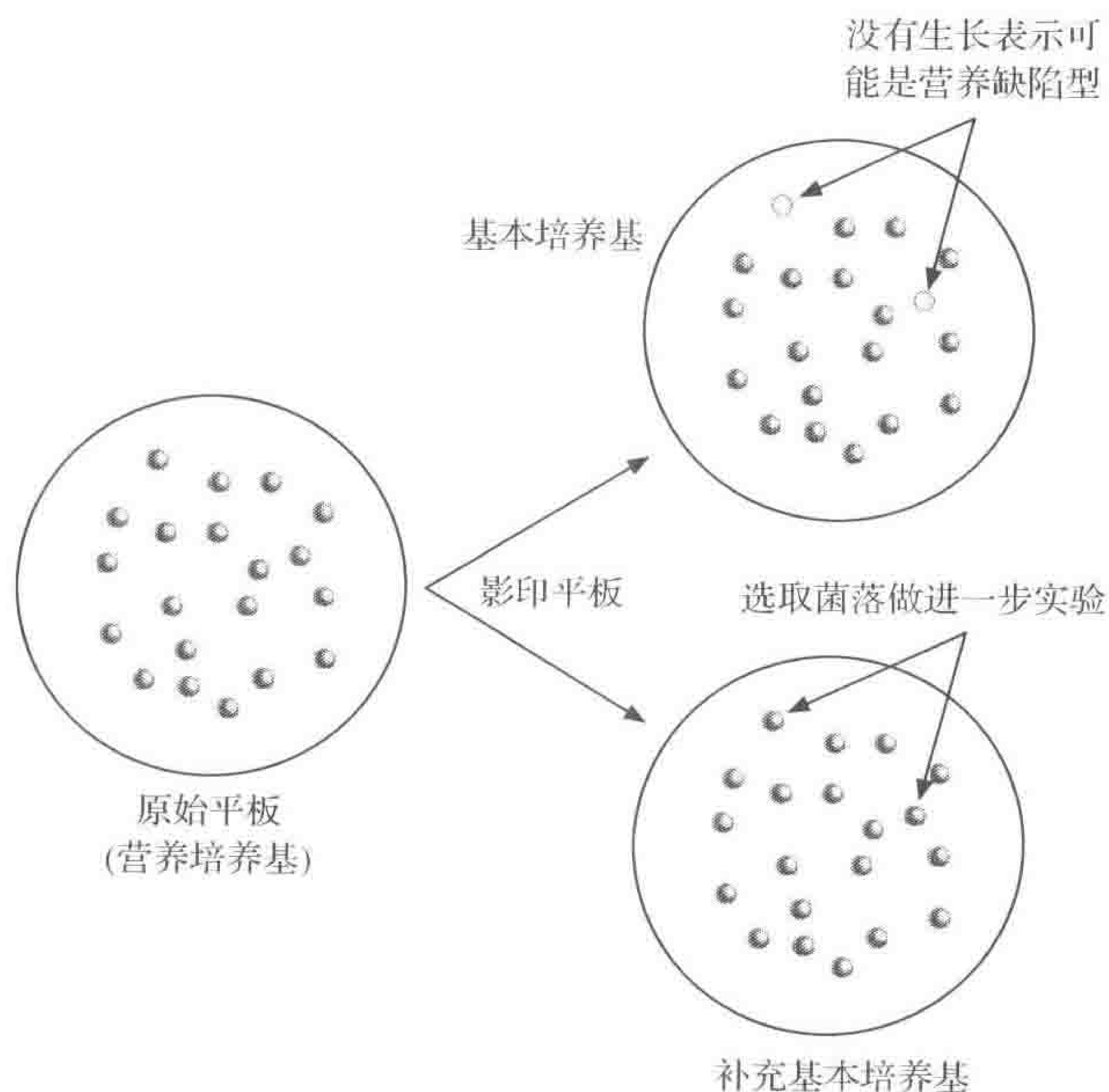


图 2.15 影印平板法分离营养缺陷型突变株

亲本菌株在两个平板上都会生长，其他类型的营养缺陷型突变株在两个平板上都不会生长。第二个平板可以淘汰不需要的其他类型的营养缺陷型突变株，也可以确保在第一个平板上不生长的菌落不是由于包有丝绒衬垫没将其影印。

即使用影印平板的方法，能处理的菌落数仍然有限。如果原始平板包含几百个菌落，则很难鉴定不能生长的菌落。如果突变频率为 10^{-6}（对大肠杆菌为高频率自发突变），就需要影印超过 1000 个平板。用诱变剂预处理细菌可增加突变量，就可以从一定数量的平板中将突变株分离出来。

使用青霉素可以增加营养缺陷型突变株的比例，因为青霉素优先杀死生长的细胞。将细菌悬浮于添加青霉素的基本培养基中，能在其中生长的亲本细菌被杀死，而不能生长的突变株则得以幸存。

用微阵列法也可进行突变株筛选，参考第 10 章。

2.6.3 其他类型突变株的分离

也可用类似的方法分离不能利用某种特异碳源（如乳糖）的突变株，例如，将细菌接种于分别含有乳糖或葡萄糖为碳源的平板上，比较菌落的生长能力。也可以选择其他方法进行分离。以大肠杆菌为例，细菌利用糖时会产酸，可通过在琼脂中添加 pH 指示剂来检测。不能利用乳糖的大肠杆菌突变株（Lac^- 突变株），可通过含乳糖和 pH 指示剂的基本培养基（突变株和亲本细菌都可生长）来检测。微生物学家常使用麦康凯（MacConkey）琼脂分离突变株。这种培养基含有乳糖和 pH 指示剂，Lac^+ 菌落呈红色，而 Lac^- 菌落为白色，从而可轻易地鉴别。

还有一个更为简单有效的方法，即利用生色底物从 Lac^- 菌落中鉴定 Lac^+ 克隆，也就是说一种底物被相关酶作用后，可发生易检测的颜色变化。例如，β-半乳糖苷酶可催化乳糖水解为葡萄糖和半乳糖，常用的 β-半乳糖苷酶的生色底物为 5-溴-4-氯-3-吲哚-β-D-半乳糖苷，即所熟知的 X-gal，它是根据其天然底物合成的类似物，即将半乳糖连接上一种染料。X-gal 本身无色，只有当连接的染料被 β-半乳糖苷酶水解释放出来才显示出颜色。Lac^+ 菌落在含有 X-gal 的培养基上呈蓝色，不产生 β-半乳糖苷酶的菌落则为白色。

2.6.4 分子生物学方法

本章前面涉及的概念及技术均以细菌自身的特征为基础。在后面的章节中，分子生物学方法将凸显其重要性，尤其是聚合酶链反应（PCR）、基因探针、Southern 印迹法和 DNA 测序等。目前这些技术已广泛应用于突变株的鉴定，因此在此进行简单介绍。以后章节将进一步讨论基因克隆及相关技术的影响，这些技术为操纵细菌遗传组成提供了特异的、通用的手段。

凝胶电泳

电泳技术的核心在于 DNA 分子在通过琼脂糖凝胶时（对非常小的 DNA 片段，则使用聚丙烯酰胺凝胶），能将不同分子质量的 DNA 片段进行分离（框 2.3）。在电泳时使用

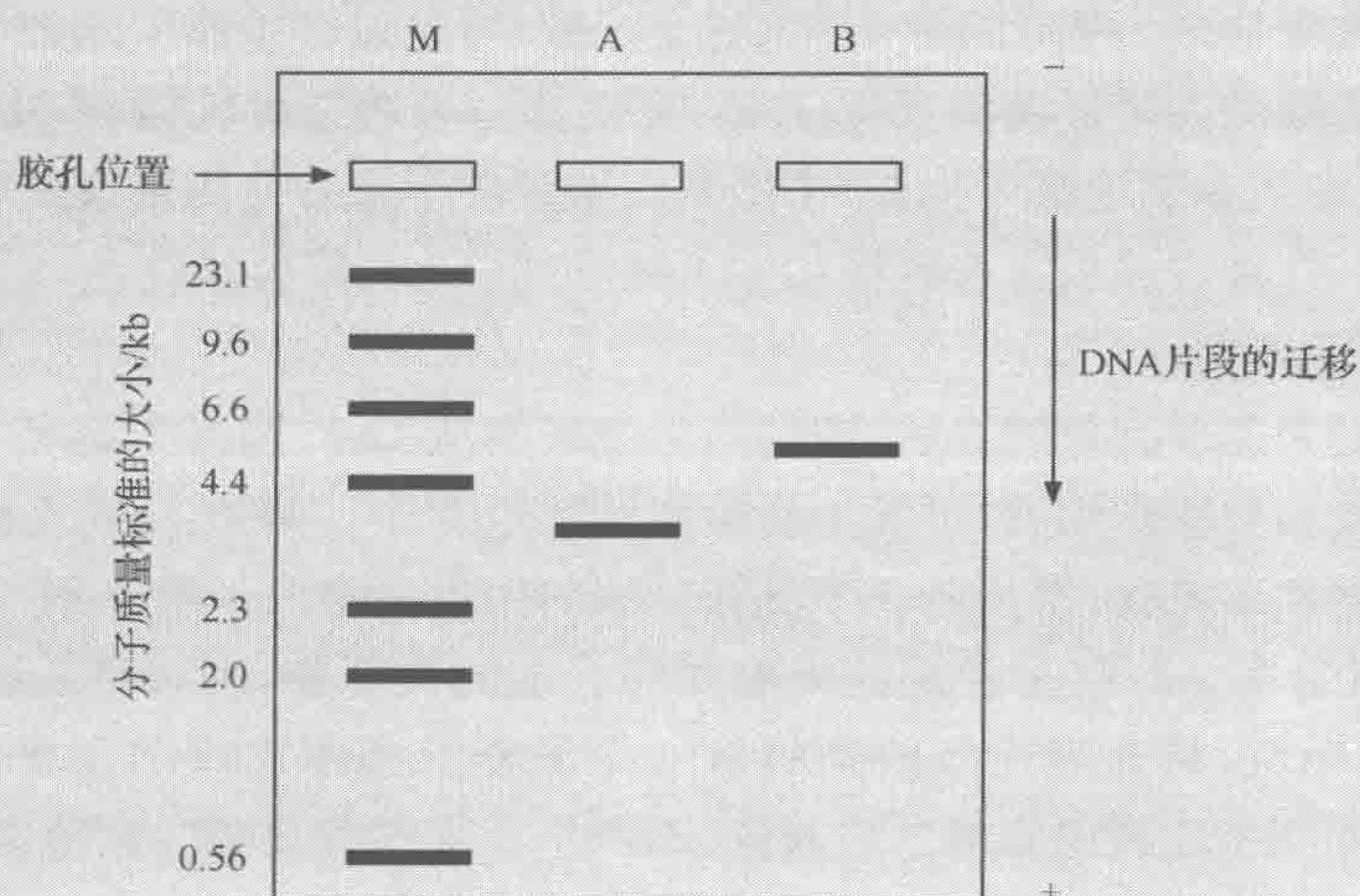
的 pH(通常为 8.3)条件下,DNA 分子带负电荷,向正极移动。小分子片段比大分子片段移动速度快,因此可用电泳方法估计 DNA 片段的分子质量。第 9 章将进一步介绍如何利用电泳检测菌株间的变异。

框 2.3 琼脂糖凝胶电泳

电泳时,通过琼脂糖凝胶(非常小的 DNA 片段,则使用聚丙烯酰胺凝胶)对不同分子质量的 DNA 片段进行分离。实验时,将样品加在凝胶的上样槽或孔内,加上电压,DNA 片段得以分离。之后凝胶用溴化乙啶染色,溴化乙啶和 DNA 形成复合物,在紫外线照射下发出荧光。

在一定条件下,DNA 片段的迁移率由其大小决定,较大的片段迁移慢。在通常使用的电泳 pH 条件下(约 8.3),DNA 分子带负电荷,基本消除了不同 DNA 片段间所带电荷的差异。可以已知 DNA 片段混合物为标准,通过分子质量对数与迁移距离之间的关系计算片段的大小。分子大小与迁移率之间的关系只有在具有同样结构的分子间才成立。因此,用线性 DNA 片段为标准,只能估计线性 DNA 的分子质量。

凝胶电泳也可用于 DNA 制备,即将目标片段从凝胶上切下,并洗脱出 DNA。



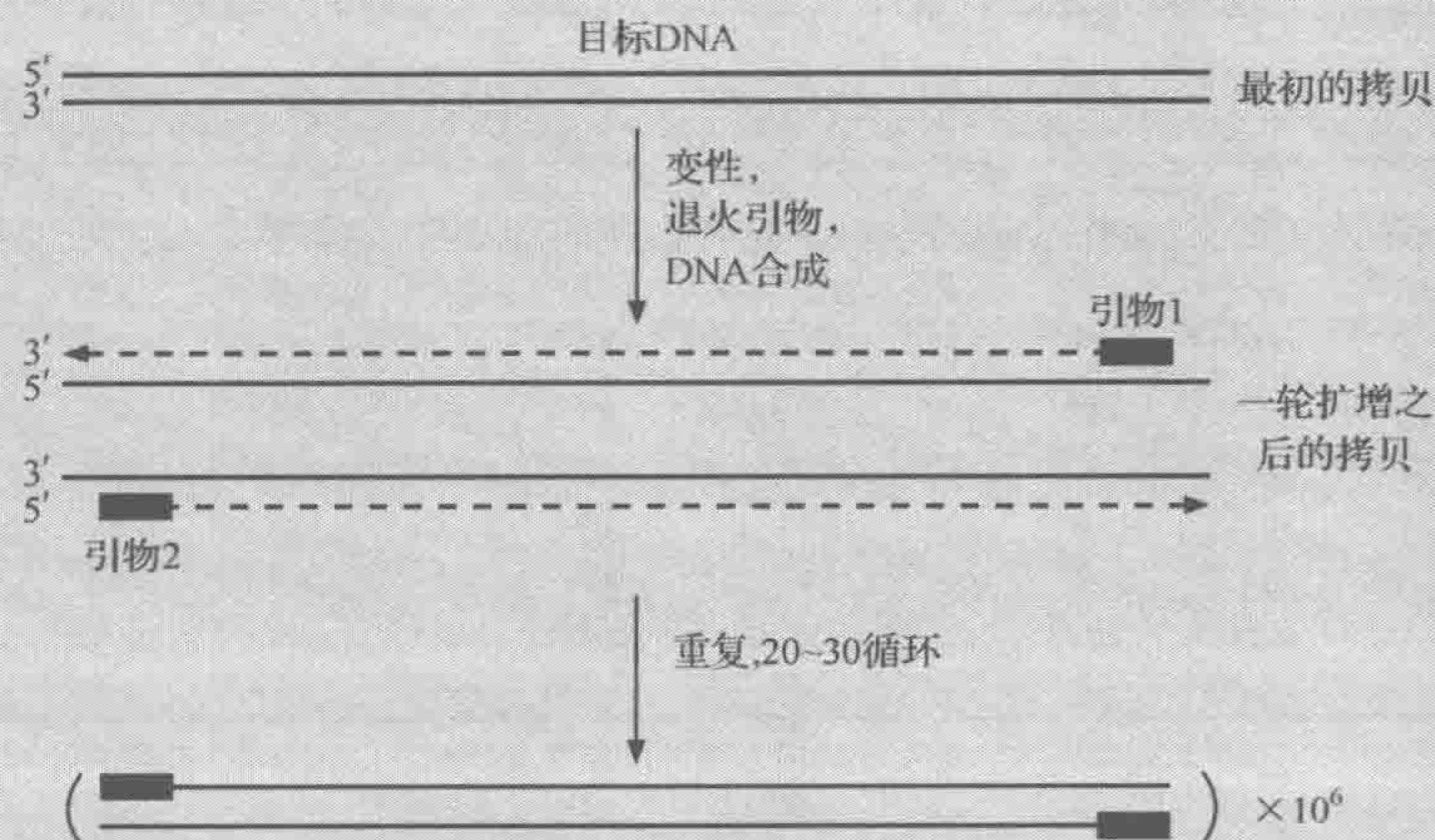
上图是用溴化乙啶染色的凝胶模拟图。泳道 M 代表标准 DNA 分子(*Hind*Ⅲ酶解的 λ DNA),分子质量以 kb(kilobases,千碱基)表示;泳道 A 和 B 为待测分子质量的 DNA 片段。

聚合酶链反应

PCR 的原理如框 2.4 所示。DNA 合成需要与模板 DNA 碱基配对的引物。用人工合成的与所研究目标基因两侧互补的寡核苷酸序列作为引物,就可以得到目标基因。加热混合物使双链 DNA 解链,引物结合到单链上,进行 DNA 的合成。重复进行上述步骤,使目标 DNA 呈指数式增加。20 个循环后,将得到百万倍的特异的目标 DNA。

框 2.4 聚合酶链反应

PCR 通过酶促反应可产生大量特异的 DNA 片段。将两条合成的能与目标 DNA 序列杂交的寡核苷酸引物与样品 DNA(模板)混合,再加入热稳定 DNA 聚合酶及 dNTP。加热混合物,使模板 DNA 变性,再冷却退火,使引物与模板结合。酶利用引物合成与每条模板链互补的新 DNA 链,使 DNA 分子成倍增加。循环不断地进行加热(变性)、冷却(退火)、合成,每经一循环,DNA 分子便成倍增加,20 个循环后,单个 DNA 分子可扩增至百万个拷贝。



聚合酶链反应用途广泛,包括以下几点。

扩增基因探针;

生产 DNA 片段用于克隆;

扩增特异的染色体片段,如用于突变体序列分析;

检测特异病原体(分子诊断);

体外遗传突变(使用改变序列的引物);

通过检测 mRNA 分析基因的表达(反转录 PCR)。

如果所研究的突变株某一特异基因长度发生显著改变(如插入或缺失),就可以通过凝胶电泳检测 PCR 产物分子质量的变化。

上述凝胶电泳只能用于检测 DNA 片段大小的改变,不能用于检测 DNA 序列的变化,但有些特殊的电泳技术可用于检测 PCR 产物间细微的差别(甚至一个碱基),这将在第 10 章进行讨论。

基因探针及 Southern 印迹法

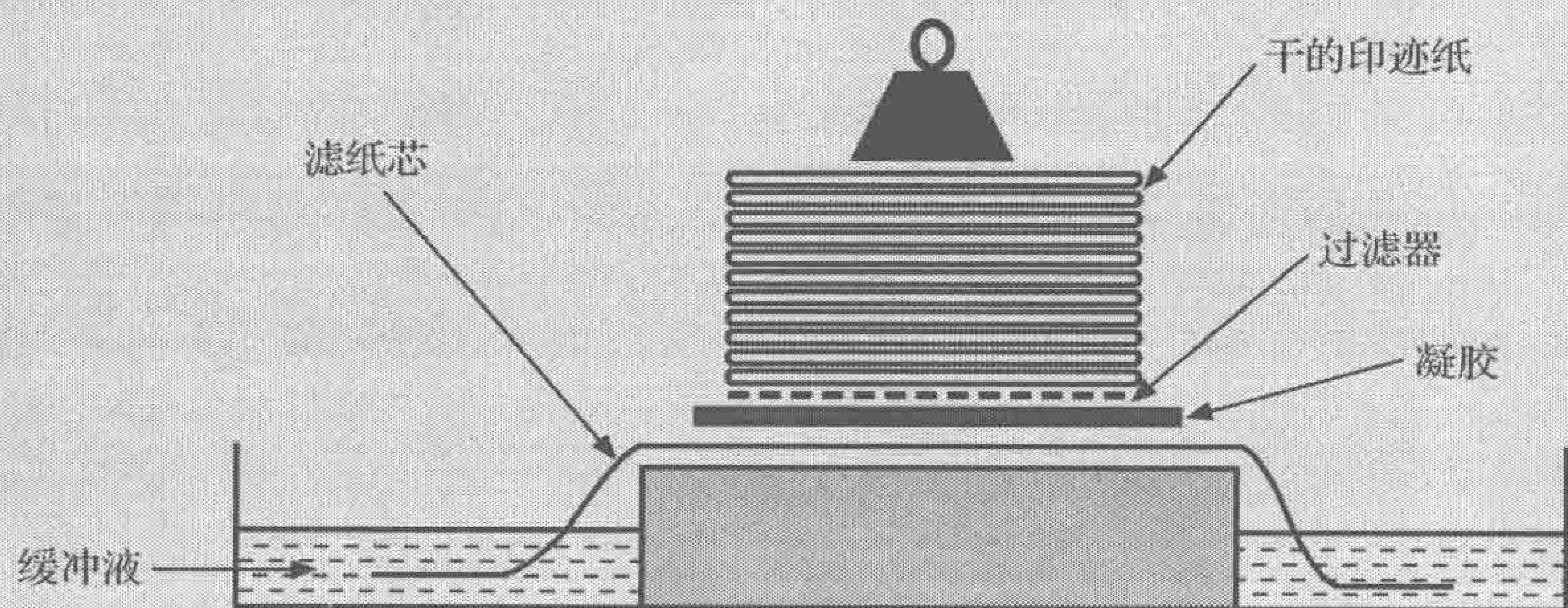
第 1 章介绍了使用标记的 DNA 片段(DNA 探针)通过杂交检测相似的 DNA 序列。如果用短的寡核苷酸作为探针,在严格条件下[高温和(或)低离子强度],可以检测出基因序列的微小变化,哪怕是单核苷酸取代。

Southern印迹法是用基因探针来检测基因结构之间，以及PCR产物之间相同和差异之处的常用技术。如框2.5所示，通过琼脂糖凝胶电泳将DNA片段进行分离，然后转移至滤膜上，用标记的探针与之杂交。通过这项技术不仅可以确认是否获得了正确的PCR产物，还可以利用高特异性的探针来检测序列之间的差异。重要的是，Southern印迹法还可以在完整基因组水解所获得的复杂DNA混合片段中检测到特异基因，因为只有目标基因才能被探针检测到。这也为该方法广泛用于分析细菌变异奠定了基础，详细内容参考第9章。

框2.5 Southern印迹法

Southern印迹法综合了凝胶电泳的强大分离能力与DNA杂交的特异性。该名字无任何地理学含义，仅以第一个发明者Ed Southern的名字命名。

下图显示了该技术的原型。在已进行DNA片段分离的凝胶上放上滤膜，再放一摞干的印迹纸，杂交纸通过毛细管作用使缓冲液通过凝胶与滤膜。DNA片段随缓冲液转移至滤膜，再用标记的DNA探针与滤膜杂交，通过放射自显影检测。



DNA测序

如果想知道所研究基因究竟发生了怎样的变化，则需要测定基因的DNA序列。这是分子生物学实验室常用的方法，不仅适用于分析个别基因或片段，还适用于分析有机体完整的基因组DNA。该方法及对细菌研究的影响将在第10章进行详细讨论。

3 基因的表达调控

适应性是细菌在不同环境条件下能否生存的关键特征。大多数生物具有的大量遗传信息、编码机制使其能应对各种挑战。依据基因组序列数据能够估测不同细菌的基因总量:在迄今已知的序列中,基因数为700～9000个,不过在某一时间点,起作用的基因只有600～800个。根据基因组序列信息,可以发现在调控基因表达中发挥作用的基因。按照ATP的消耗量计算,基因表达的代价是高昂的(合成一个蛋白质需消耗约3000个ATP分子),因此,为了防止由不必要物质的合成所引起的浪费,基因表达需要精确调控。以在不同环境下生存的铜绿假单胞菌(*Pseudomonas aeruginosa*)为例,它需要468种蛋白质来调控其对不同刺激的应激反应,而特异性地寄生于人胃中的幽门螺旋杆菌(*Helicobacter pylori*)只有18种。

除了应对环境或其他刺激以调控基因表达外,细菌细胞合成结构蛋白、核糖体蛋白等蛋白质的水平较高。而如一些调节蛋白等其他蛋白质,则维持在较低水平。在针对环境变化或处在不同的生长阶段时,虽然这些蛋白质的表达水平可能升高或降低,但是其最高表达潜力是固定的。幸运的是,固定及可变调控的机制是相似的。因此,可以一并考虑。

从基因的结构到作为其最终产物的酶的活性的信息流向图(图3.1)中,可以发现控制主要发生在3个关键阶段,即mRNA的转录、将mRNA翻译为蛋白质及该蛋白质酶活性的控制。在此框架下,许多潜在的调节因素如下。

1. 基因拷贝数。大体上讲,如果一个基因有几个拷贝,该基因表达的水平就可能比较高,但这一关系并不是必然的线性关系。

2. 主要由RNA聚合酶控制的转录起始水平(启动子活性)决定的基因转录效率,不论是固定的或可变的控制,对细菌而言,这是影响单个基因表达的主要因素。

3. mRNA的稳定性。重要的是,要认识到特定mRNA的量是由产生速度及每个分子在细胞内保持活性状态的时间跨度等综合因素决定的。许多细菌的mRNA寿命很短,半衰期约为2min。细菌mRNA的不稳定性是其对环境变化反应快速的一个关键特征。然而,一些种类的mRNA较其他种类的更稳定,在某些情况下,其半衰期长达25min。rRNA和tRNA等其他形式的RNA则更稳定,这归因于这些分子具有更高程度的二级结构。

4. mRNA翻译成蛋白质的效率。启动效率(与核糖体结合)及其他影响翻译速度(主要是密码子选择)的因素决定了这一效率。

5. 蛋白质产物的稳定性。相对于mRNA,蛋白质的量反映了其合成的速度及其稳定性。不同的蛋白质稳定性差别很大,这与其所发挥的不同功能有关。形成细胞结构某一部分的蛋白质要比传递信号以开启瞬时细胞活动的蛋白质更稳定。

6. 翻译后效应。这包括许多事件,例如,将多肽转换成具有生物活性构象过程中所

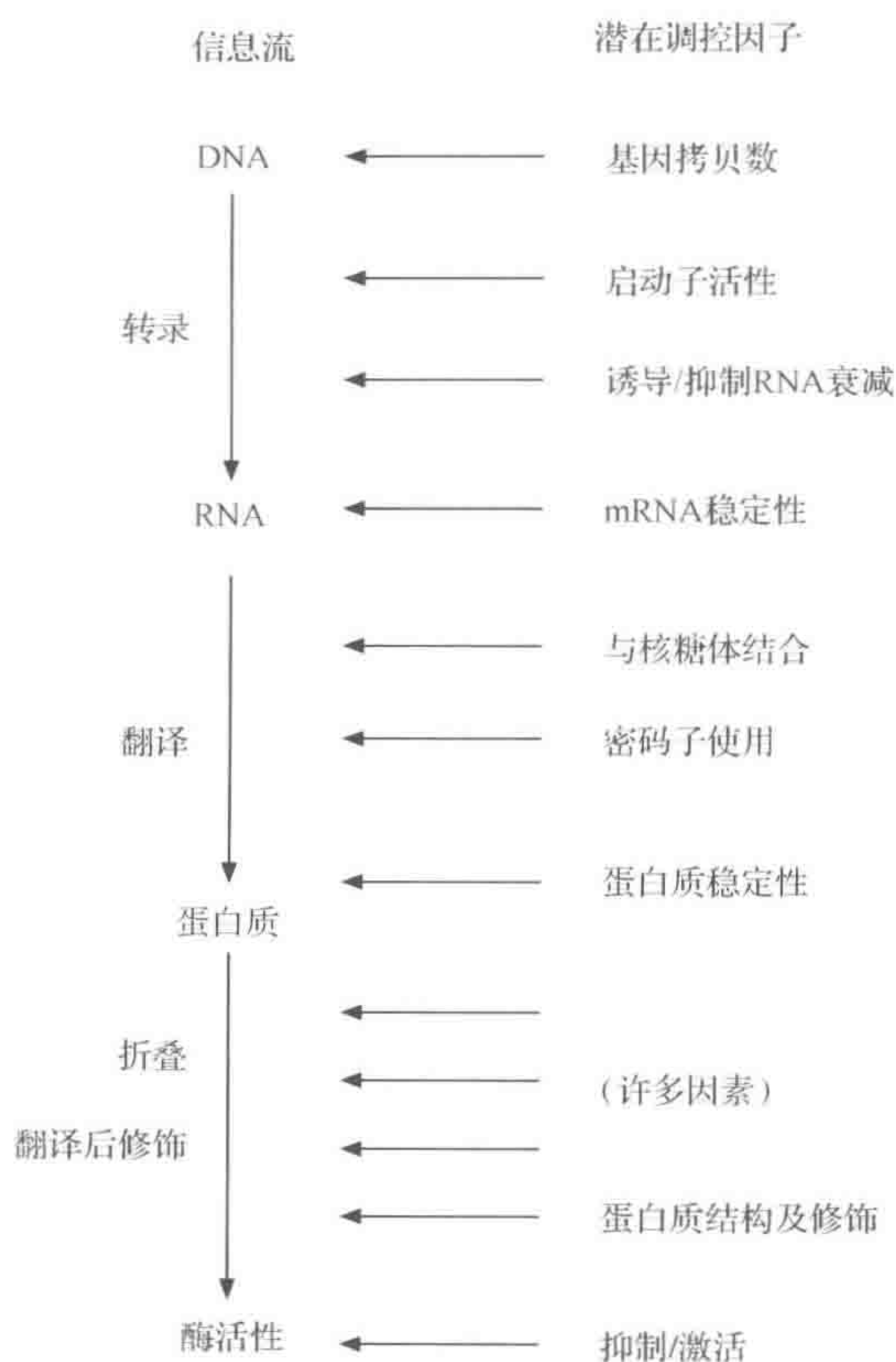


图 3.1 信息流向及调控因子

必需的蛋白质折叠，以及影响蛋白质活性的共价修饰。作为特定蛋白质功能的调节机制，磷酸化尤其重要。另外，细胞代谢及生理学的控制受临时因素的影响，包括个别酶的抑制和激活等。

现在可以更详细地阐述这些因素。

3.1 基因拷贝数

细菌染色体上的许多基因是单拷贝的(但也有例外，如合成 rRNA 的基因)。因此，基因拷贝数不是调控许多细菌细胞正常代谢活动的重要手段。然而，当需要质粒介导的某些特征时，尤其是异源 DNA 的克隆和表达，基因拷贝数是非常重要的。一些质粒在细菌细胞中具有很高的拷贝数，可高达上千拷贝数(见第 5 章)，其携带基因的高水平表达反映了这一特性。

3.2 转录控制

3.2.1 启动子

调控细菌基因表达的主要手段是调节该基因转录 mRNA 的量。它主要是由 RNA

聚合酶与启动子结合的亲和力所决定的(见第1章)。强启动子效率高,可促进高水平转录,而弱启动子则导致低水平的转录。因此,启动子的性质很重要,它决定了不同基因的潜在表达水平。

根据对数以百计的启动子区的比较分析结果,得到了共有序列(图3.2)。许多大肠杆菌启动子含有两个重要序列,它们参与RNA聚合酶的识别,这些类似于TTGACA和TATAAT的序列分别位于转录起始位点上游35和10个碱基的位置,因此,分别被称为−35和−10位点,后者也被称为Pribnow盒。每2s能够介导基因转录的强启动子与这一理想的保守序列有相同或相近的序列,而弱启动子的这些区域发生了碱基替换或两个基序间距不同。因此,这些基因大约每10min转录一次。

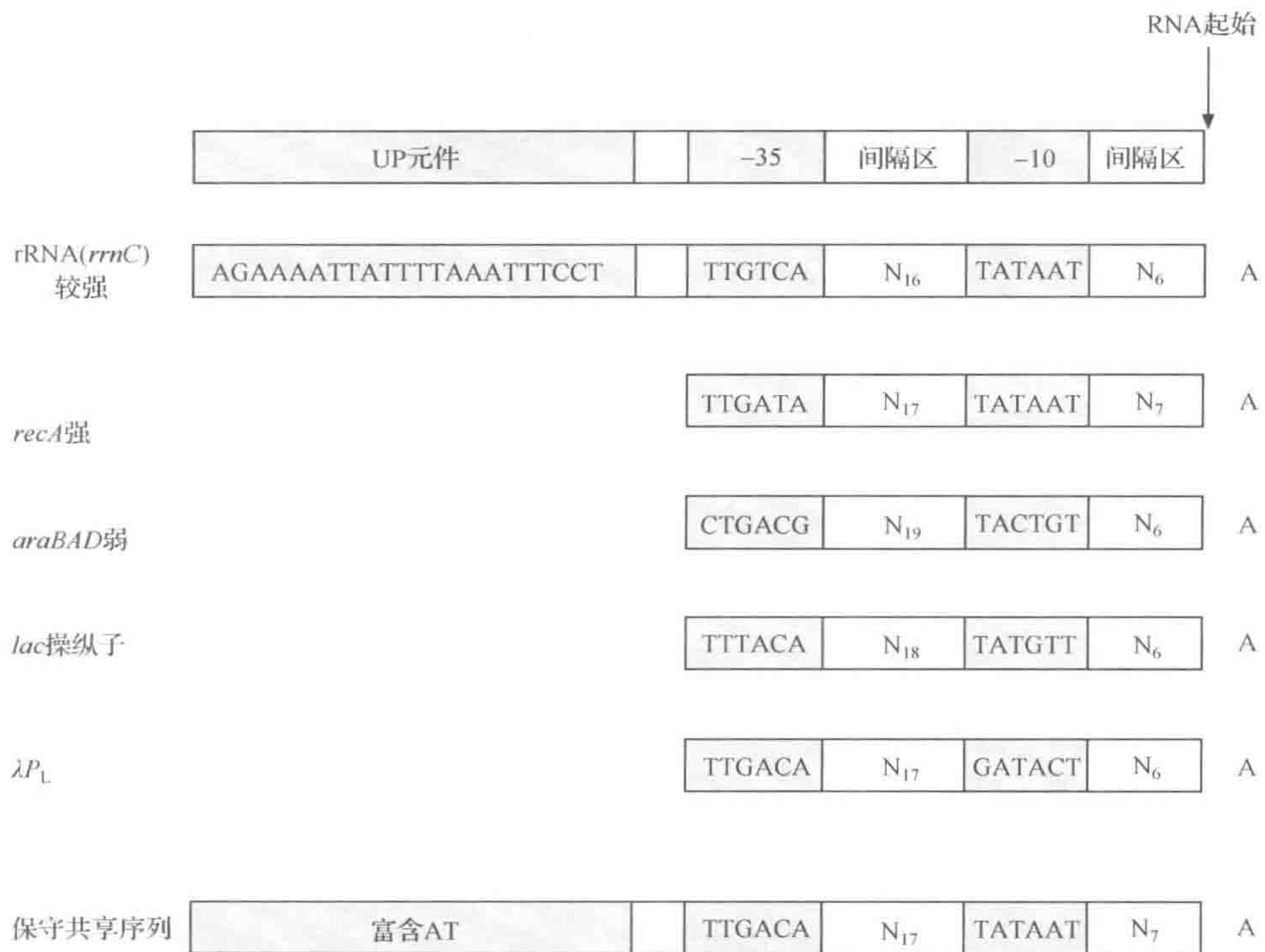


图3.2　典型的大肠杆菌启动子结构

尽管酶识别位点需要特异序列是可以理解的,但限制两个基序间距的理由可能比较难以理解。与启动子区域相比,RNA聚合酶很大,理论上几个碱基不应该影响RNA聚合酶与两个基序的结合。然而,因为双链形成双螺旋结构,增加(或减少)两个基序间距意味着它们不再处于DNA的同侧(图3.3),所以RNA聚合酶将不能同时与两个基序结合。

−35和−10基序之外的启动子序列也可能影响它的效率。指导rRNA转录的操纵子是大肠杆菌最活跃的启动子,其驱动的转录约占细胞内全部转录的60%,比其他约2000个启动子指导更多的mRNA转录。除−10和−35基序外,启动子还有一个富含AT序列的第三种元件,被称为UP元件(图3.2),它可提高转录活性30~90倍,并且使

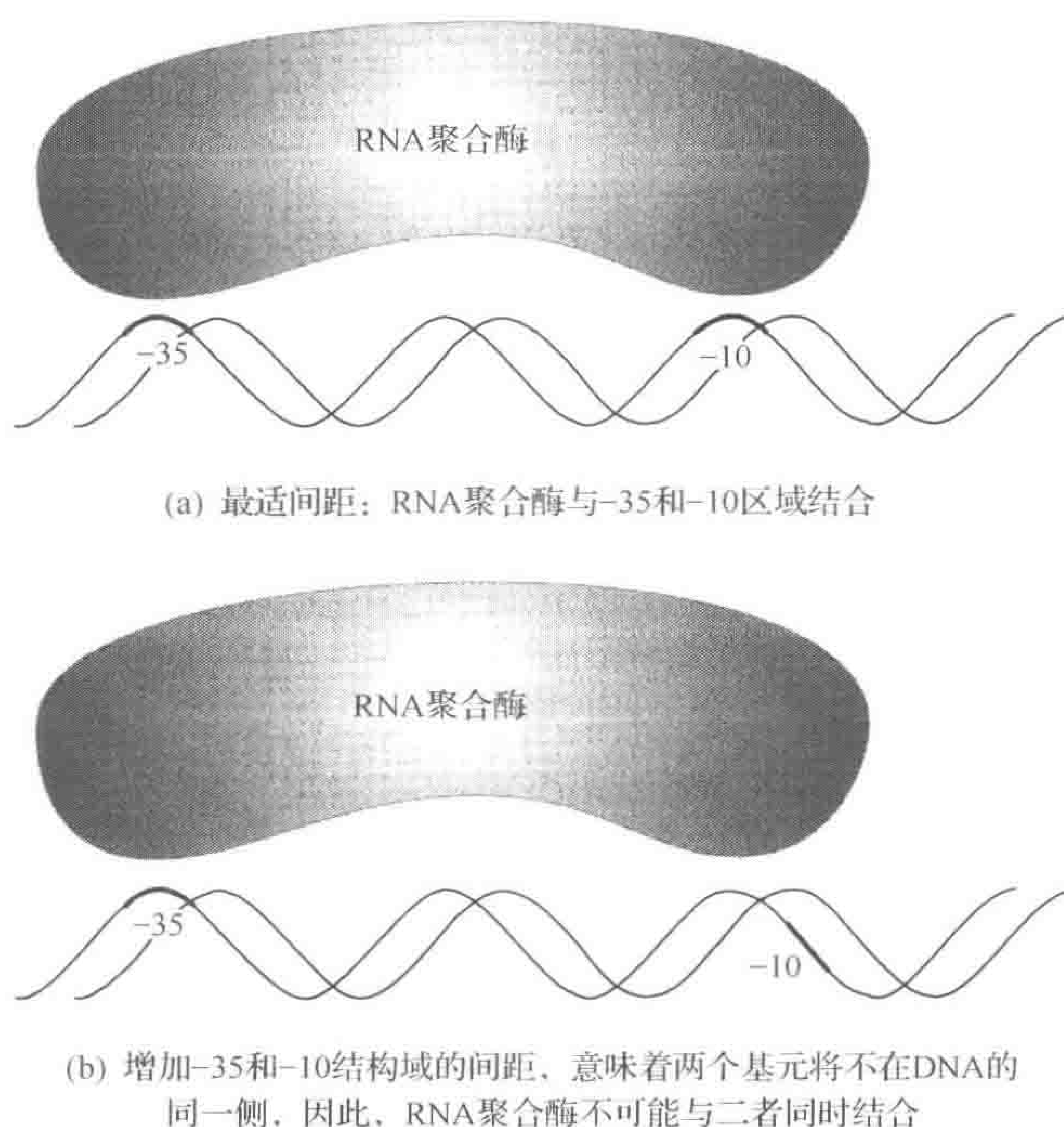


图 3.3 启动子－35 区和－10 区间距离的重要性

启动子变得更强。而且正如在后面将要讨论的那样，DNA 局部结构将影响启动子活性，尤其是 DNA 超螺旋及折叠将影响 DNA 聚合酶打开螺旋的速度进而决定转录是否能够顺利进行。

值得注意的是，这里讨论的启动子是在“正常”状态下使用的。在特殊条件下，如热休克，可能利用替代性启动子，这些启动子与上述的－10 和－35 基序相似度很低甚至完全不匹配(见后续章节)。

操纵子和调节子

在第 1 章中提到，一些细菌基因可能是单一转录单位的一部分，即操纵子。从一个启动子开始，整个操纵子被转录成一个长的 mRNA 分子，其编码的蛋白质单独进行翻译。因此，转录调控能够将操纵子作为一个整体进行，其结果是协调地调控操纵子内的基因表达，这种基因调控方式是细菌所特有的。例如，大肠杆菌基因组有 4289 个基因，但是它们分属于 578 个已知操纵子。

大肠杆菌 *lac* 操纵子是最为熟悉的一个例子(图 3.4)。它含有结构基因 *lacZ*(编码β-半乳糖苷酶)，*lacY*(编码通透酶，用于乳糖的摄入)和 *lacA*(编码硫代半乳糖苷转乙酰酶)。这些基因的 5′端存在一个包含启动子及部分重叠的操纵基因调控区。这一操纵基因和分开的编码抑制蛋白的 *lacI* 基因在操纵子的诱导能力中相互关联，这将在本章的后续内容中进行介绍。

在某些情况下，由一个调节两个反向启动子的操纵基因位点协同控制几个基因的表达(图 3.5)。例如，编码异亮氨酸和缬氨酸生物合成中所需的一个酶的 *ilvC* 基因和编码

调控蛋白的 *ilvY* 基因是反向转录的，即反向基因，但其转录由一个操纵基因控制。因为是一个操纵基因，所以基因被协同控制，尽管产生两个不同的 mRNA 分子，但也称为一个操纵子。操纵子的一般规律是一个操纵子基因转录成一个 mRNA，但这是一个例外。

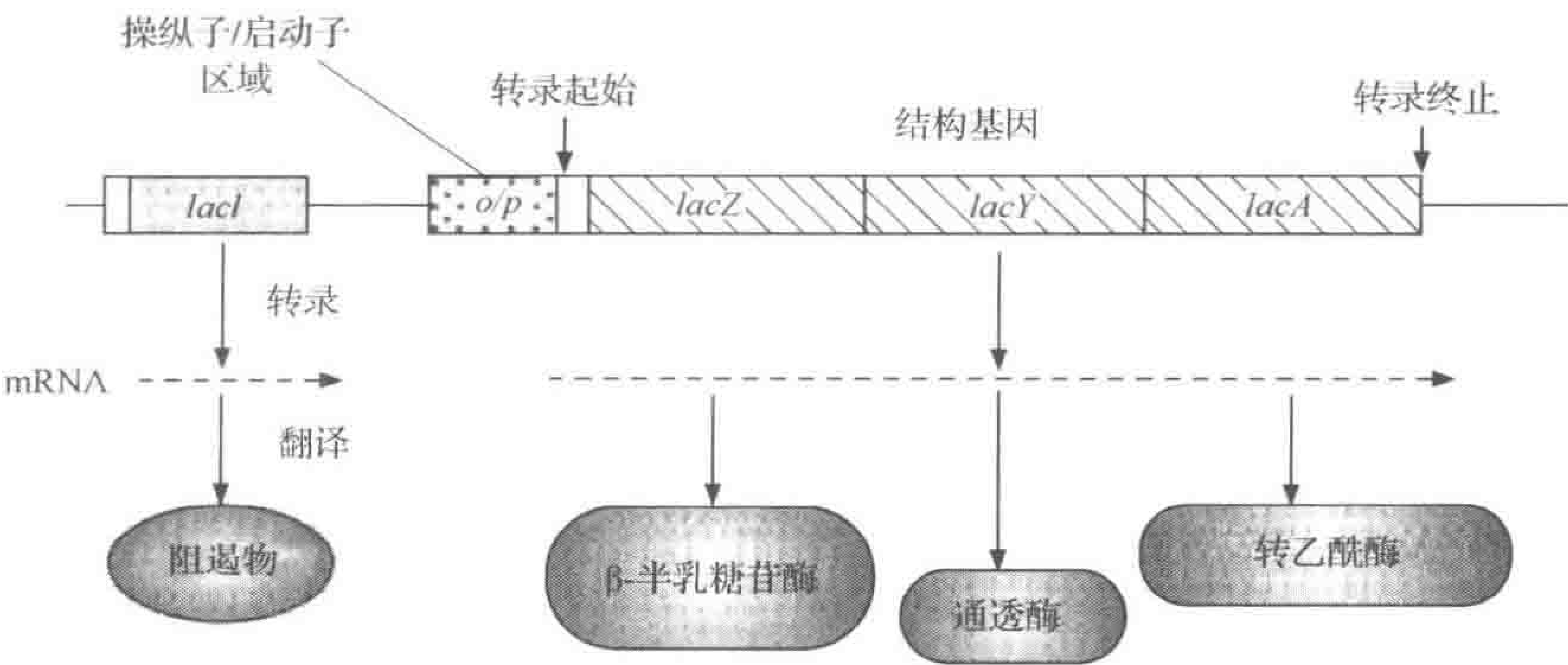
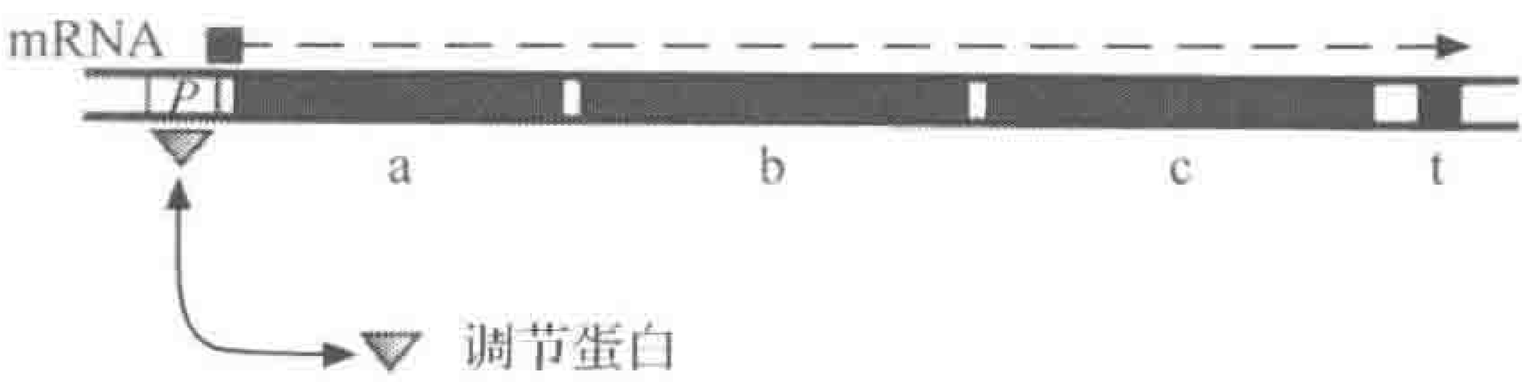
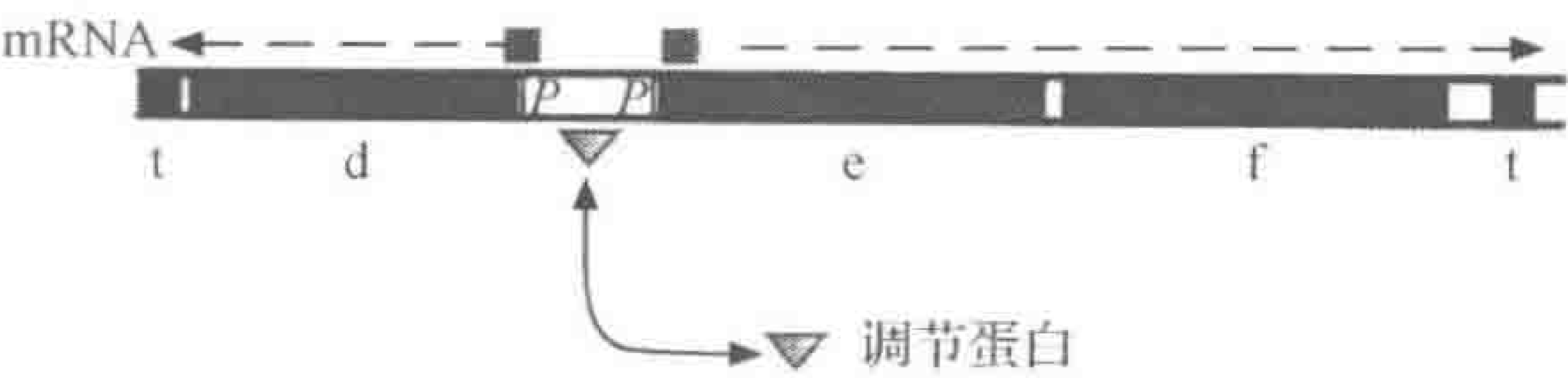


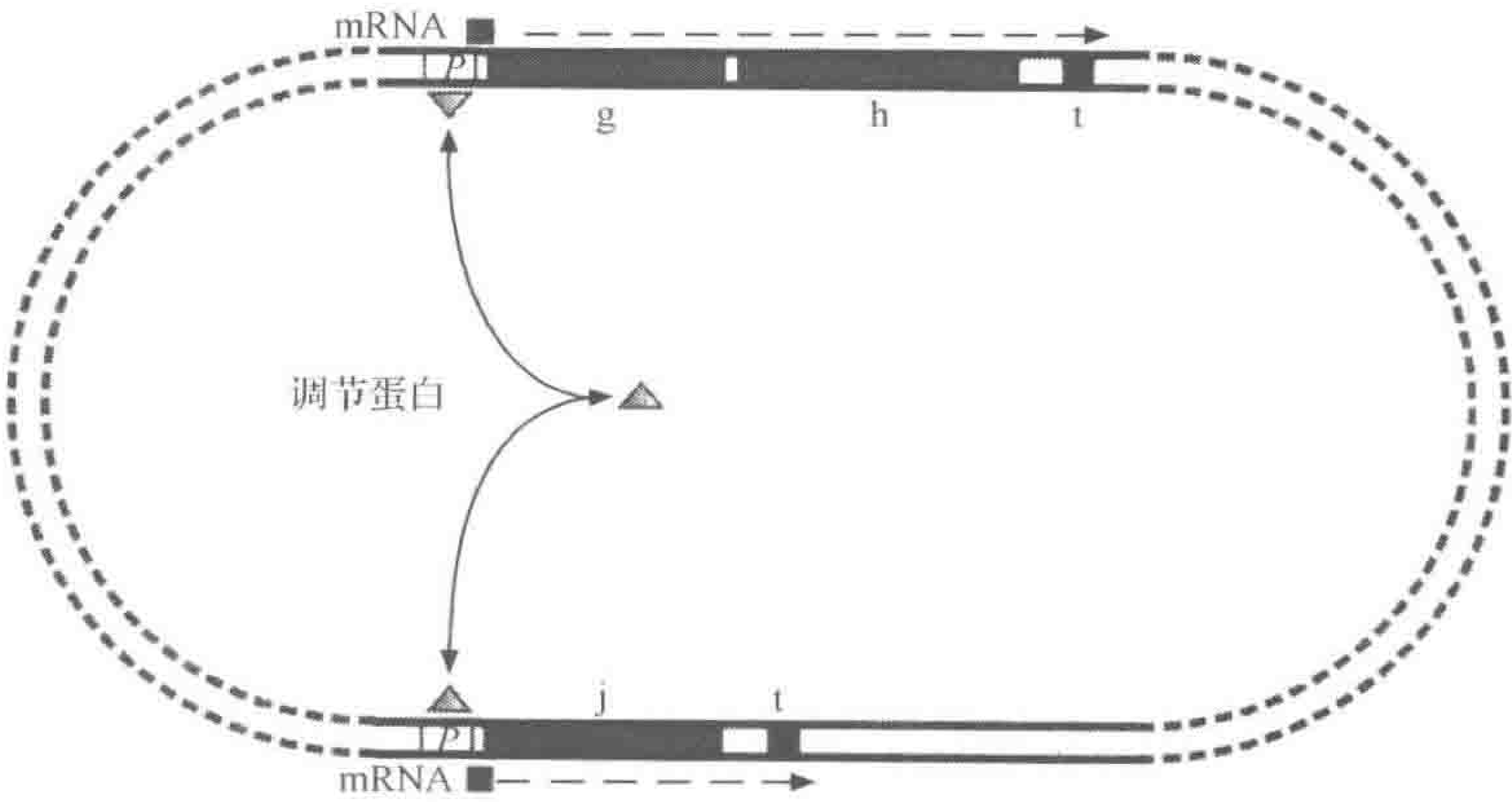
图 3.4　乳糖操纵子的结构



(a) 操纵子1：一个操纵子/启动子控制一套相邻基因



(b) 操纵子2：一个操纵子控制一个逆向启动子



(c) 调控蛋白与几个操纵子结合，控制基因组上不同部分的基因

图 3.5　操纵子和调节子的结构

并非所有协同控制基因的排列均按操纵子的方式进行。在某些情况下，染色体上不

同部位的基因组合是以协作模式被调控的。这样的一组基因或操纵子，从不同的分开的启动子位点表达，但同时受同一调节分子控制，被称为调节子（图 3.5）。例如，精氨酸的生物合成需要 8 个基因（*argA*～*argH*），但是在大肠杆菌中，只有 *argC*、*argB* 和 *argH* 3 个基因组成具有一个启动子的操纵子。第四个基因 *argE* 是由邻近的启动子控制的反向转录（这是前面提到的反向操纵子的又一例证）。而其他的 3 个基因（*argA*、*argF* 和 *argG*）则位于染色体的不同位置，并具有自己的启动子。尽管精氨酸生物合成基因散在分布，但它们协调表达。

选择性启动子和 σ 因子

在第 1 章中提到，RNA 聚合酶由 5 个亚基组成，$\alpha_2\beta\beta'$ 4 个亚基构成其核心蛋白，第五个可分离的亚基是 σ 因子。σ 因子负责聚合酶特异性地识别启动子区域的两个保守的核苷酸序列，所以，σ 因子决定酶的特异性。

σ 因子识别前面提到的启动子保守序列，因为在大肠杆菌中其分子质量约为 70kDa，通常也称作 σ^{70}。该亚基负责识别处于指数生长期细胞所必需的用于多数基因转录的启动子。因为这些基因是细胞周期和包括糖酵解、三羧酸循环及 DNA 复制等基础代谢所必不可少的，所以，有时也成为“管家基因”。然而，大多数种类的细菌有几个 σ 因子。用不同亚基替换基本的 σ 因子会根本性地改变 RNA 聚合酶识别的启动子序列（图 3.6）。这些可选择性的 σ 因子使细菌基因表达发生整体的改变，以便更好地应对特殊的环境应激。例如，气温的突然升高可激发许多细菌表达可抵抗高温危害作用的热休克蛋白。大约 40 个热休克基因的启动子仅被含有 σ^{32} 的 RNA 聚合酶识别（也称为热休克 σ 因子），通常，它的半衰期很短，然而，一旦存在于高温环境下它不能被降解，可激活靶基因的转录。因此，σ 因子使细菌表达一组基因，其产物是应对热应激所必需的。因为许多这类基因散在分布于基因组中，它们形成了热应激调节子系统。图 3.6 中列举了其他 σ 因子及其调节活性和识别序列。

-35区域	间隔区	-10区域		σ因子	调节基因
TTGACA	$N_{16\sim18}$	TATAAT		σ^{70}	传统管家基因
CTTGAAA	$N_{13\sim15}$	CCCATNT		σ^{32}	热休克基因
CTGGNA	N_6	TTGCA		σ^{54}	同化氮的基因
		CTACACT		σ^{38}	静止状态/常规应激基因
GTTTAA	$N_{12\sim14}$	GGGTAT		σ^{B}	常规应激基因*
TAAA	N_{15}	GCCGATAA		σ^{28}	鞭毛基因

* 为枯草芽孢杆菌，其他都是大肠杆菌

图 3.6 可选择性(可变)σ 因子和启动子的结构

细菌所含有的σ因子的数量差别很大。例如，研究很透彻的大肠杆菌和枯草芽孢杆菌两个生物模型分别有7个和18个σ因子，框3.1中列举了大肠杆菌中σ因子所具有的不同的调节作用。枯草芽孢杆菌含有额外的11个σ因子，形成这种现象的主要原因之一是在这些额外的专用的σ因子调控下，能够完成复杂的孢子形成过程（见本章的后续部分内容）。因为σ因子同时调控基因表达的诱导或抑制，所以细菌所拥有的σ因子越多，它的能力也越多样化。目前已经研究的数百种细菌基因组中的证据支持了这一普遍趋势，如对幽门螺旋杆菌和铜绿假单胞菌的σ因子进行比较，前者可导致人消化道溃疡并且很少在其特定的胃壁外发现它们，而后者可在水、土壤、航空燃料、化妆品、排水沟等很大范围的环境中存在，同时可随机地感染人类、昆虫和植物。因此，幽门螺旋杆菌仅有3个σ因子，而铜绿假单胞菌有24个σ因子。天蓝色链霉菌（*Streptomyces coelicolor*）主要生存在土壤环境中，但产生不同的有价值的二级代谢产物，其中部分产物是抗生素，并且它具有很复杂的形态学分化的生命周期。因此，它需要精细的调节系统，事实上，它有65个σ因子来完成其复杂的基因调控。

框3.1　7种不同大肠杆菌σ因子的功能

σ因子	调节基因的功能	调节基因的数量
RpoD/σ^{70}	常见的管家	≈1000
RpoH/σ^{32}	热休克	≈40
RpoN/σ^{54}	氮素同化	≈15
RpoS/σ^{38}	静止期/一般胁迫	≈100
RpoF/σ^{28}	趋化性/鞭毛合成	≈40
RpoE/σ^{24}	**周质蛋白，表面蛋白**	**≈5**
FecI/σ^{18}	**柠檬酸铁运输**	**≈5**

黑体文字部分表示的是σ因子的胞质外家族

根据序列分析结果，σ因子可分为σ^{54}和σ^{70}两大组。σ^{70}组是主要的因子，还可继续分为亚组。最大化最多样化的σ因子群对介于外膜或外周胞质空间的蛋白质的表达进行调控，故命名为胞质外功能（ECF）σ因子。许多细菌，尤其是含有较大基因组的细菌，含有多种ECF σ因子，而且这些调控者往往较其他类型的σ因子的总和还多。例如，铜绿假单胞菌有24个σ因子，其中19个属于该亚群。因为它们经常调控细菌表面分子的表达，而这些分子是细菌和人类免疫系统接触的第一位点，所以它们对细菌的毒力影响很大。

当枯草芽孢杆菌受到刺激时，其生存机制之一是形成孢子，产生具有强抵抗力的孢子内壁。这是一个复杂的转录调节过程，包括调控100多个基因的表达及6个不同σ因子的按次序地介入，图3.7是一个简单的流程。当开始处于营养缺乏之时，主要σ因子σ^{A}和低丰度因子σ^{H}指导一组基因转录，其产物可使膜发生不对称内陷，然后与细胞分离形成前芽孢。在此过程中，染色体各有一个拷贝分配到两个子细胞中。另一个σ因子σ^{F}出现在隔膜形成前，由于抗σ因子的存在，其没有活性（见后续内容）。当隔膜形成时，σ^{F}变得有活性，并且只在前芽孢中，它促进新的一组基因的表达。随后，孢子形成的第三个特异

σ因子 σ^E 也开始有活性，但只是在母细胞中。因此母细胞中不同的基因被激活，前芽孢被第二层膜吞噬。然而，由孢子染色体浓缩依赖于另一个σ因子 σ^G 介导的基因表达，该σ因子在发生吞噬的前芽孢中变得有活性。σ^G 所驱动基因的表达导致前芽孢发生了改变，导致母细胞中另一个σ因子 σ^K 的激活。在这一时期中，基因的表达产物参与孢子形成的最后阶段，包括皮质和表皮层的形成，并将孢子包裹其中，这种结构对干燥及消毒剂等化学试剂具有强抵抗力。

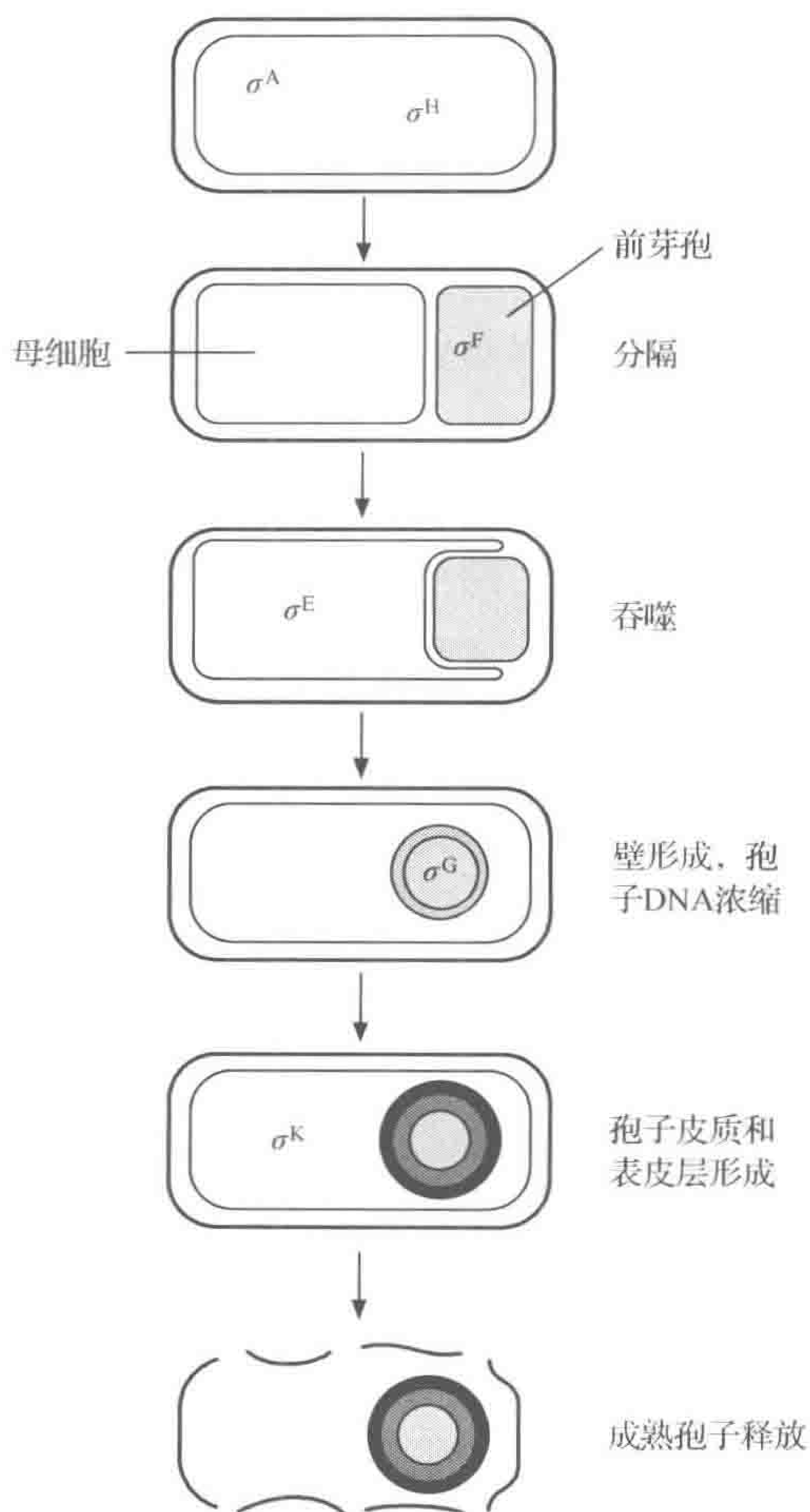


图 3.7　枯草芽孢杆菌的孢子化

图示该过程中每个σ因子首先被激活

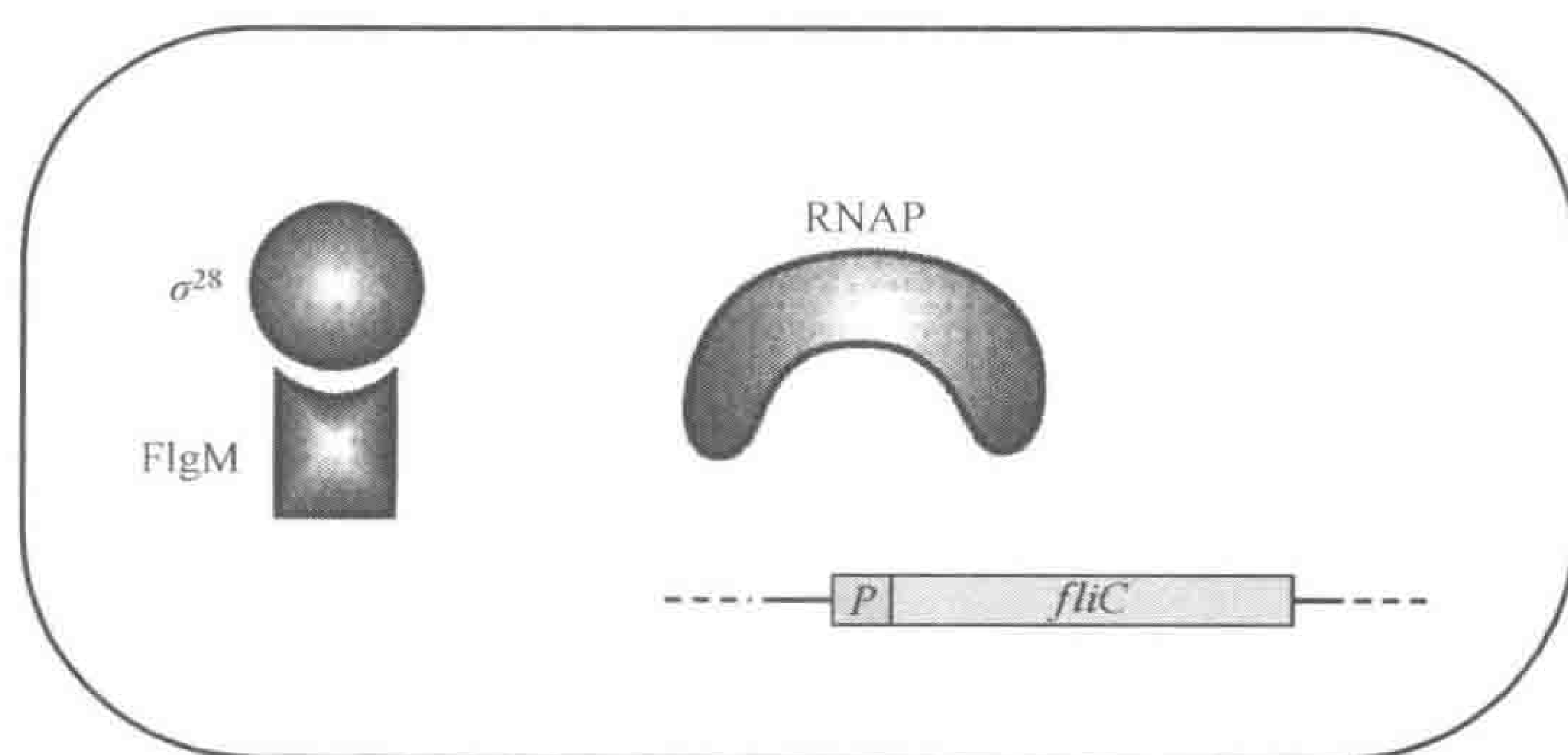
在这一过程中，最重要和有趣的是，被间隔开的细胞的两部分必须平行发展。母细胞的每个阶段与前芽孢的对应阶段相匹配，反之亦然。如果不同步，则不能形成孢子。在每个发育阶段，两部分之间存在一种被称为交互调制的信号转导形式，它刺激下一个阶段的出现。这是细菌发育调控的最容易理解的例子。

以下将进一步列举一些σ因子的例子，在第 4 章中，将特别介绍由噬菌体所携带的σ因子。

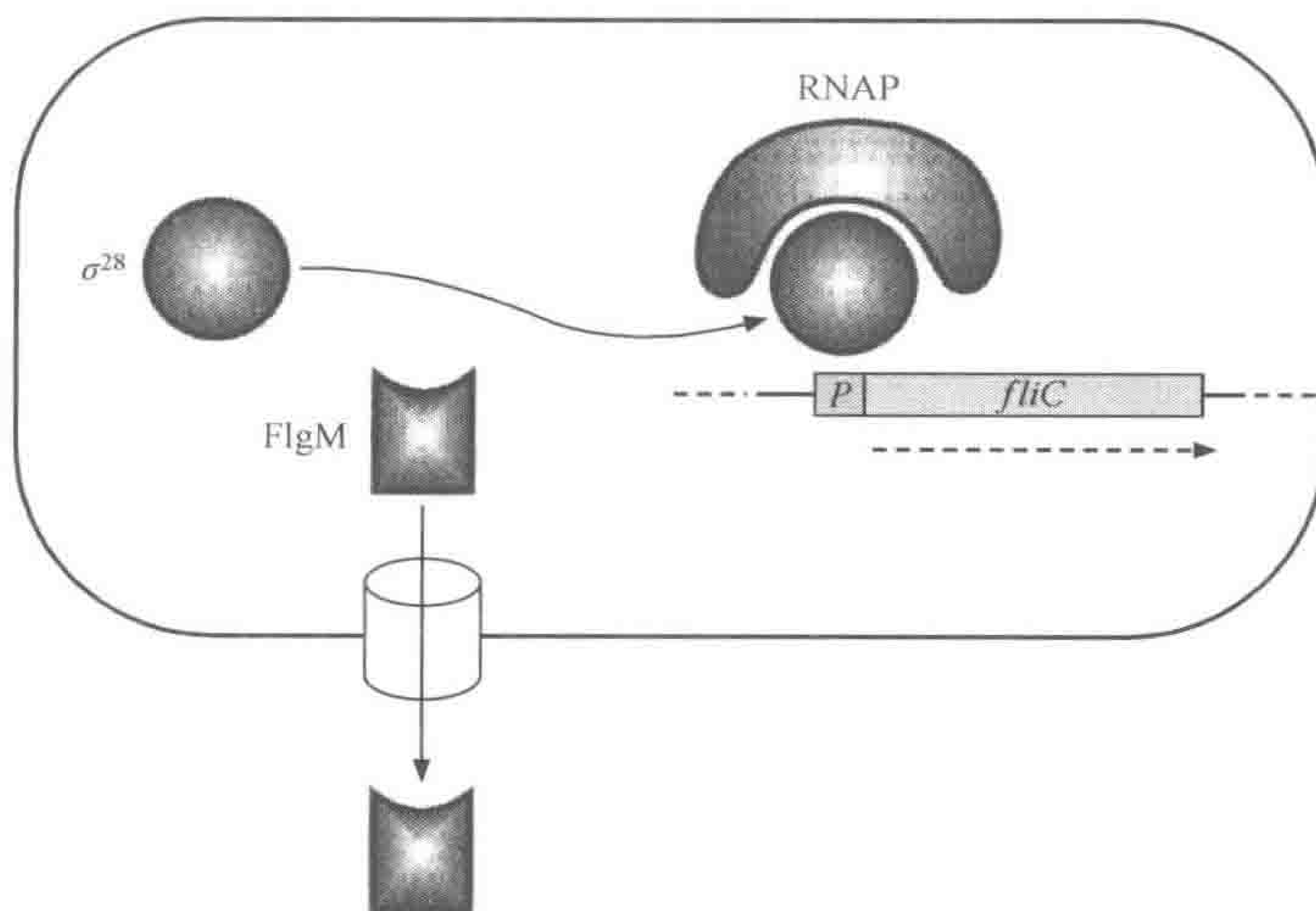
抗 σ 因子

在不同刺激的应激反应中，σ 因子激活基因的能力需要控制，以便应激条件不存在时，它们能失去活性。以前面提到的 σ^F 因子为例，抗 σ 因子是完成这一过程的有效方式之一，它们可以与特异性的 σ 因子结合并阻止其与核心 RNA 聚合酶相互作用。前面提到的 ECF σ 因子是这一机制很好的例证，它们常与一个具有抗 σ 因子的跨膜蛋白共转录，后者与前者结合并抑制 σ 因子活性。当细菌感受到来自环境的刺激时，会释放 σ 因子，后者与 RNA 聚合酶结合刺激转录。这些特别成对的抗 σ 因子/ECF σ 因子与后面所描述的双组分调控系统中是相似的，原因是它们都将细胞质内的转录反应与信号连接到一起，该信号是由包膜蛋白的包外域感知的。

抗 σ 因子也存在另一种转录调控机制，伤寒沙门氏菌鞭毛的生物合成采用了这一模式(图 3.8)。鞭毛是细菌细胞的最复杂的结构之一，需要超过 52 个基因的协调表达，占



(a) 在钩状基体形成前，FlgM与σ^{28}结合并阻断其与RNA聚合酶的结合



(b) 钩状基体形成后，FlgM由基体形成的孔中排出，缺少FlgM时，σ^{28}能够与RNA聚合酶结合并且启动鞭毛细丝蛋白FliC的表达

图 3.8　抗 σ 因子 FlgM 调控鞭毛细丝蛋白 Flic 的合成

基因组的1.5%,鞭毛特异性的σ因子σ^{28}是主要调控者之一。细菌鞭毛的装配依序进行,首先是每层亚结构,然后是外层结构。因此,细菌面临的问题之一是如何感知内部亚结构的建成,进而启动与之连接的外层细丝的产生。抗σ因子FlgM的调控机制便是一个很好的例子。在内层亚结构完成前,FlgM与σ^{28}结合并抑制其功能,因此细丝不能合成。然而,一旦鞭毛的细胞内部分钩状基体形成,其成为空的通道和输出通道,不仅输出细丝蛋白,而且包括FlgM。因此在钩状基体形成的同时,FlgM被输出细胞,缺失FlgM后,σ^{28}可激活细胞细丝基因的表达。

3.2.2 终止子,衰减子及反终止子

在第1章中介绍了位于操纵子末端终止转录的转录终止子的结构和功能。转录终止也发挥着调控操纵子内不同基因表达水平的作用。

一个启动子控制一个操纵子内基因的转录,因此,它们转录的启动或终止是同时进行的。然而,尽管细胞同时需要这些产物,但并不是所有产物必须处于同一水平。转录终止提供了解决这一问题的一种方式(另外一种方式是稍后要描述的翻译调控)。

如图3.9所示,基因*a*和*b*的产物要比*c*和*d*的水平高。位于*b*和*c*之间的终止位点*t*1使RNA聚合酶停顿。然而,只有部分转录终止,而其余的转录将重新开始。这导致部分和全长的mRNA分子共存。此过程称作衰减,与此相关但有明显不同的*trp*操纵子的衰减过程将在本章的后续部分中介绍。

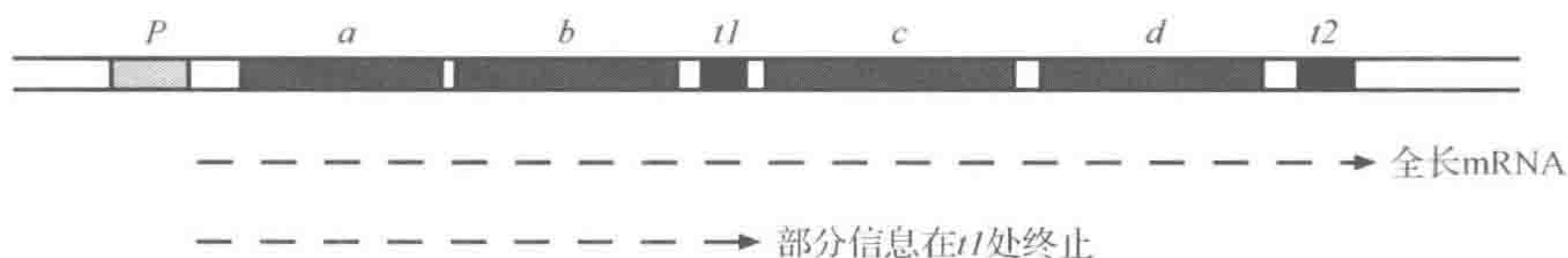


图3.9 操纵子内的衰减作用

操纵子内存在弱的转录终止位点(*t1*)导致末端基因(*c*和*d*)减少表达;强终止子*t2*引起全长mRNA末端终止

在某些情况下,抗终止子蛋白对RNA聚合酶进行修饰,使之忽略相关的终止子,因此转录不能终止。抗终止最好的例证是噬菌体λ的调控(见第4章)。

3.2.3 诱导和抑制:调节蛋白

乳糖操纵子

细胞启动或关闭转录的重要手段之一是利用阻遏蛋白和操纵子序列结合,后者与启动子相邻或者重叠。阻遏蛋白与操纵基因结合阻碍转录起始。以*lac*操纵子为例,与*lac*操纵子分开的*lacI*基因编码阻遏蛋白(见图3.4)。与乳糖阻遏蛋白结合的操纵子序列,与启动子重叠并且其前面大约20个碱基被转录(图3.10)。不是所有操纵子和启动子位点的相对位置都是相同的,但最终结果是一样的,阻遏蛋白与操纵子序列的结合阻止了RNA聚合酶接触启动子或妨碍转录起始。

*lac*阻遏物是由4个相同的亚基组成的多聚体,其二级结构含两个α螺旋,两个α螺

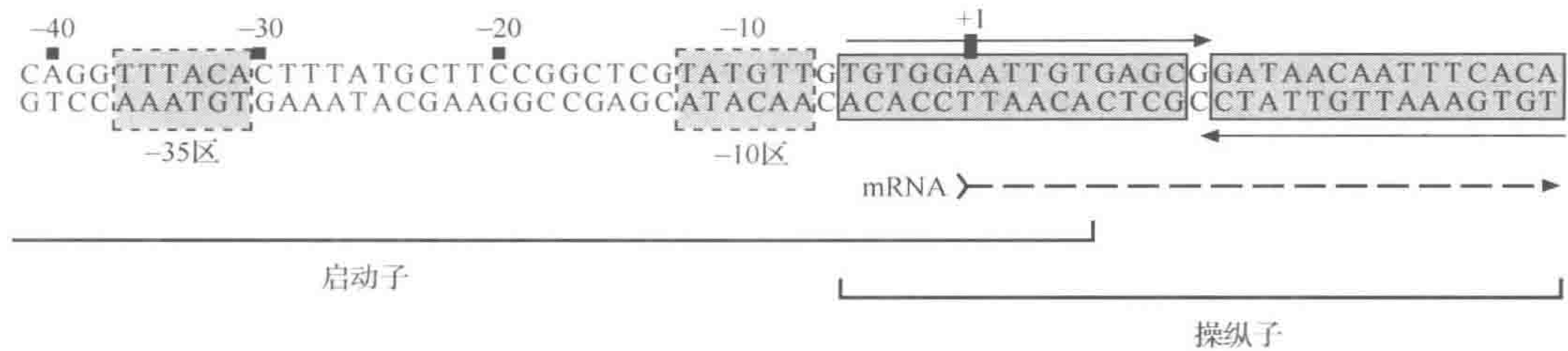


图 3.10 乳糖操纵子的操纵子/启动子区域的结构

旋之间由几个氨基酸连接，使之彼此之间形成固定的角度，该构象是已知的螺旋-转角-螺旋基序，是 DNA 结合蛋白的特征，它可以使部分蛋白质与 DNA 主槽相吻合并与 DNA 操纵子序列特异性地接触。乳糖操纵子序列成对对称，换言之，如图 3.10 表中文字所示，部分序列是反向重复的(不是很精确的)。如图 3.11 所示，调节蛋白的四聚体结构中的每半部分可识别结合位点的一半序列，因此，它增加了操纵子的特异性和亲和力(图 3.11)。

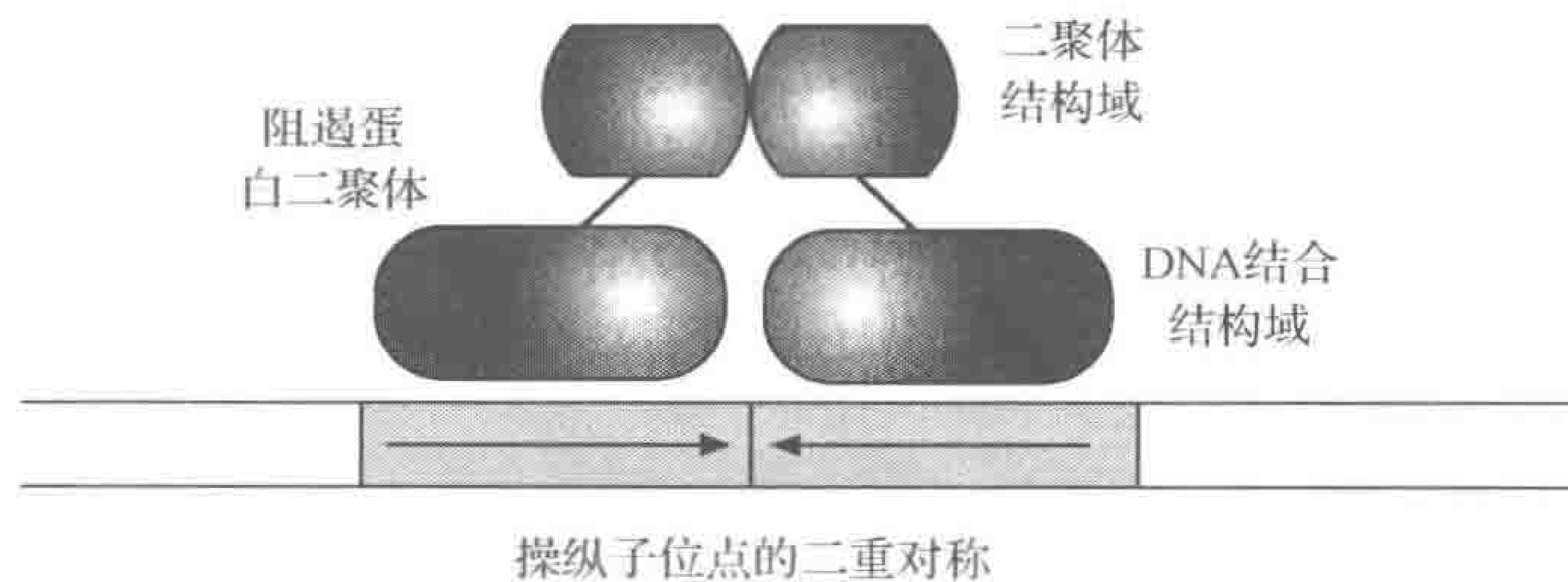


图 3.11 调控蛋白的二聚体与操纵基因成对称位点的结合

乳糖阻遏蛋白与异乳糖也有亲和力，后者是乳糖的衍生物。异乳糖与阻遏蛋白的结合使其蛋白质构象发生变化，导致其不再与操纵子位点结合。因此，当乳糖缺乏时，阻遏蛋白激活，与操纵子结合并阻止乳糖操纵子的转录(图 3.12)。当乳糖存在时，阻遏蛋白失活，操纵子表达，细胞产生乳糖代谢所需要的酶。

在自然状态下，大肠杆菌的乳糖发酵是诱导表达的过程，即只有在乳糖存在的条件下表达。在实验室条件下，乳糖不是使用方便的诱导物，因为它一旦诱导表达就开始代谢，所以势必削弱它的诱导效应。然而，其作为诱导物及底物的特性是分开的，前者依赖于阻遏蛋白的识别，而后者是通过 β-半乳糖苷酶识别和启动的。因此，可以设计乳糖类似物作为诱导物，因为它们可与阻遏蛋白结合，但不能被 β-半乳糖苷酶降解。合成类似物异丙基-β-D-硫代吡喃半乳糖苷(IPTG)便是这种安慰诱导物。

反之亦然，一些化合物是 β-半乳糖苷酶的底物，而不是诱导物，因为它们不能被阻遏蛋白识别。X-gal 是众所周知的底物而不是诱导物，它是生色底物，当它被 β-半乳糖苷酶水解后产生蓝色。因此，为了获得蓝色克隆，必须在培养基中加入诱导剂 IPTG 等。

乳糖操纵子的调节变异

上述模型中的很多证据来源于对乳糖操纵子的调节变异研究，其可分成如下几个类别。

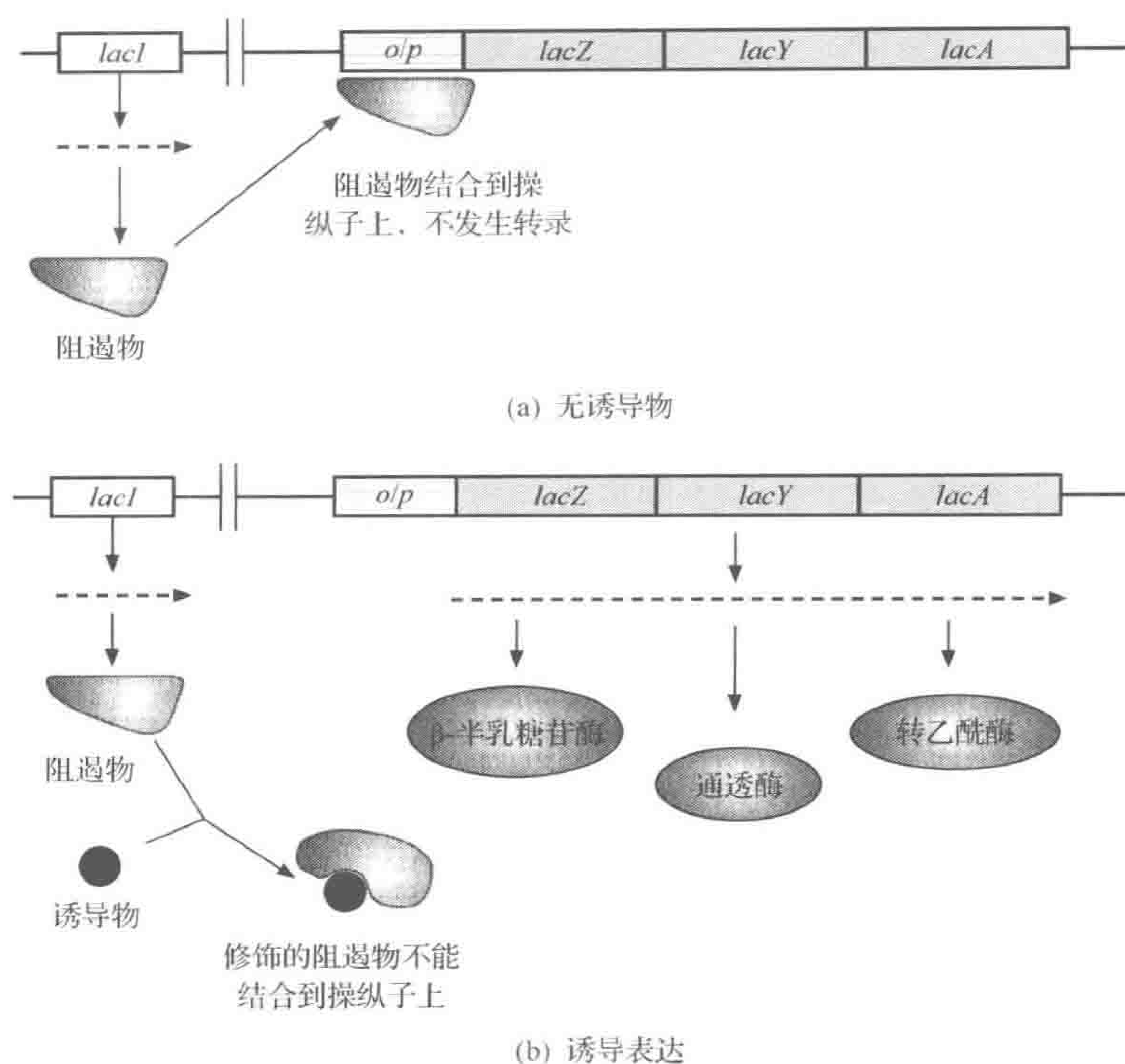


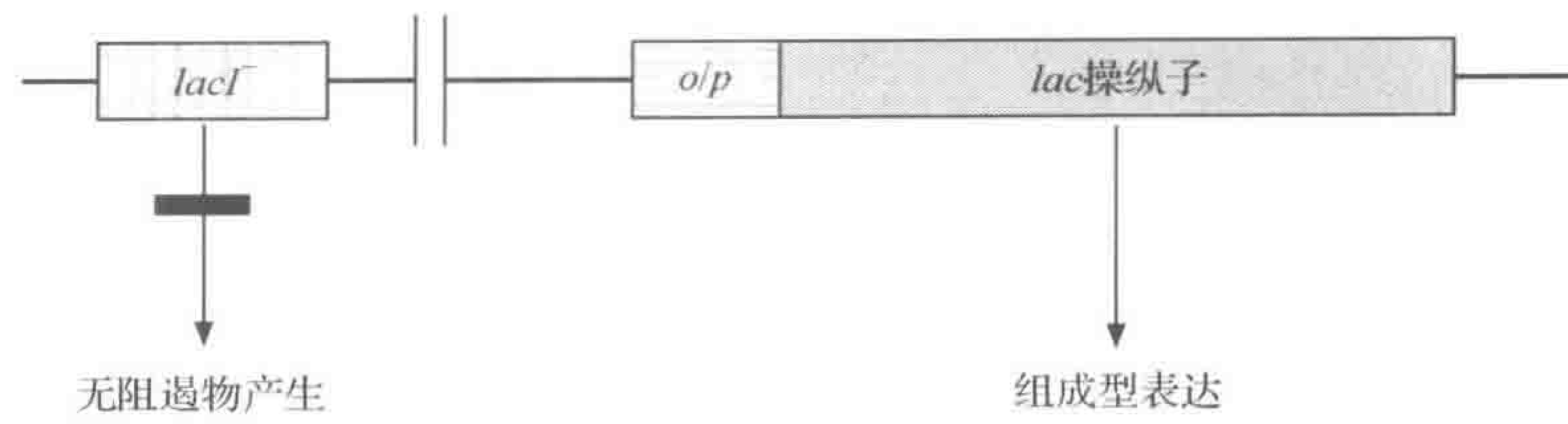
图 3.12　乳糖操纵子的调节

1. 组成型突变。即使在缺失诱导物的情况下，乳糖操纵子的酶在细胞内依然可进行最大限度的表达。这类突变分为两种，一种是 *lacI* 突变，使其不能合成阻遏蛋白或阻遏蛋白不能与操纵子结合。另一种是操纵子基因组成性突变(O^c)，操纵子发生变异后，其不能被阻遏蛋白识别。可以构建部分二倍体菌株，其调控基因位于染色体上，而野生型调控基因位于 F′质粒上(见第 5 章)。当导入携带有功能的 *lacI* 基因的质粒后，$LacI^-$ 缺失突变菌株恢复了可诱导性(图 3.13)。换言之，*lacI* 突变是隐性遗传的，这与模型是吻合的，因为质粒携带的 *lacI* 基因将合成有活性的阻遏物，它能与染色体上的 *lacI* 操纵子结合(该阻遏物被认为是反式作用)。

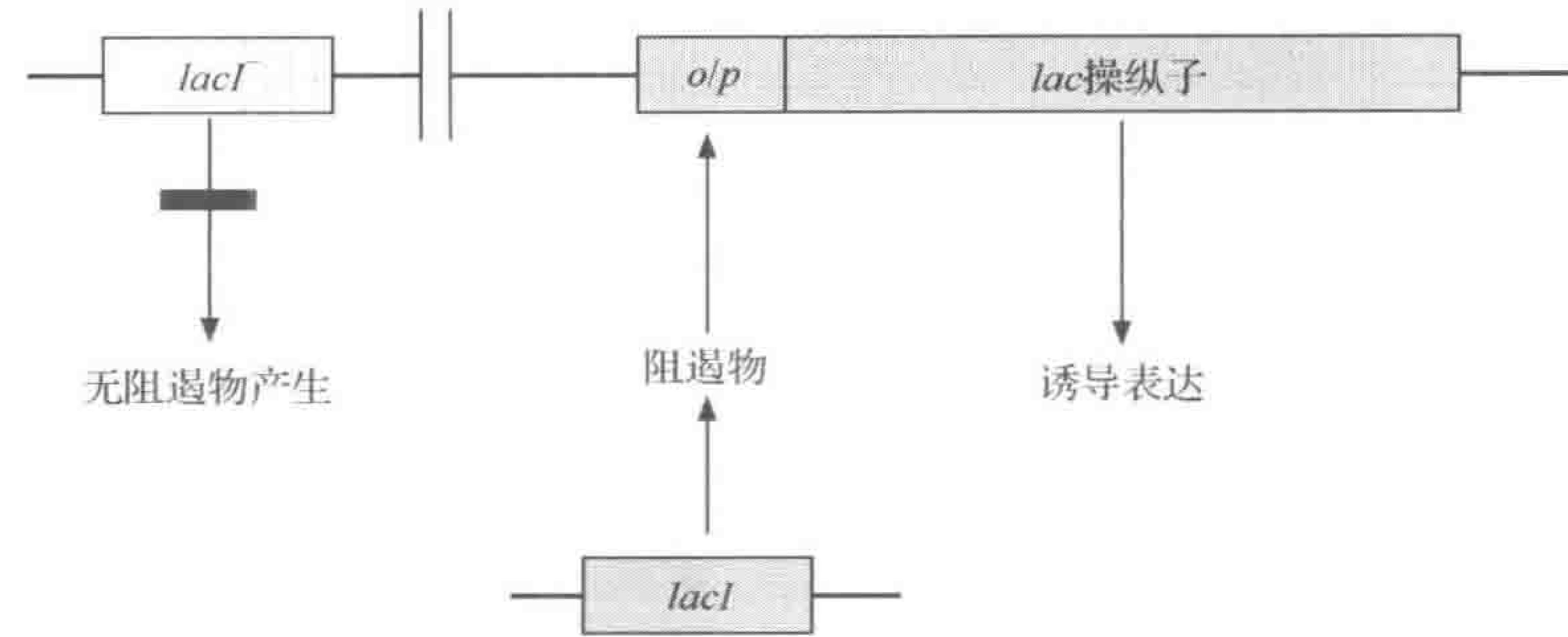
另外，在操纵子组成型突变中，质粒的介入不能恢复其诱导性(图 3.14)，由质粒编码的阻遏蛋白同样不能与染色体操纵子结合。这些突变被称为顺式显性：操纵子中的基因组成性地表达而与导入质粒上的基因无关。它们不是完全显性的，如果质粒含有一个功能性 *lac* 操纵子，而不仅仅是 *lacI* 基因，这个操纵子可以具有正常的诱导性，并且不受染色体上 O^c 突变的影响，如图 3.14 所示。因此，O^c 突变是顺式而不是反式作用。

2. 非诱导突变。这些突变不受诱导物存在或缺失的影响，但是在某种情况下，酶水平总是很低。有很多原因可以导致这种现象，但最重要的原因是不同类型的 *lacI* 突变可以减弱阻遏蛋白对诱导物识别和反应的能力。

3. 超阻遏物(*lacIq*)突变。这些细胞过表达阻遏蛋白，通常是由于 *lacI* 基因启动子的突变，值得注意的是 *lacI* 基因不属于 *lac* 操纵子，并且二者的转录有不同的启动子。当



(a) *lacI*基因突变导致组成型表达

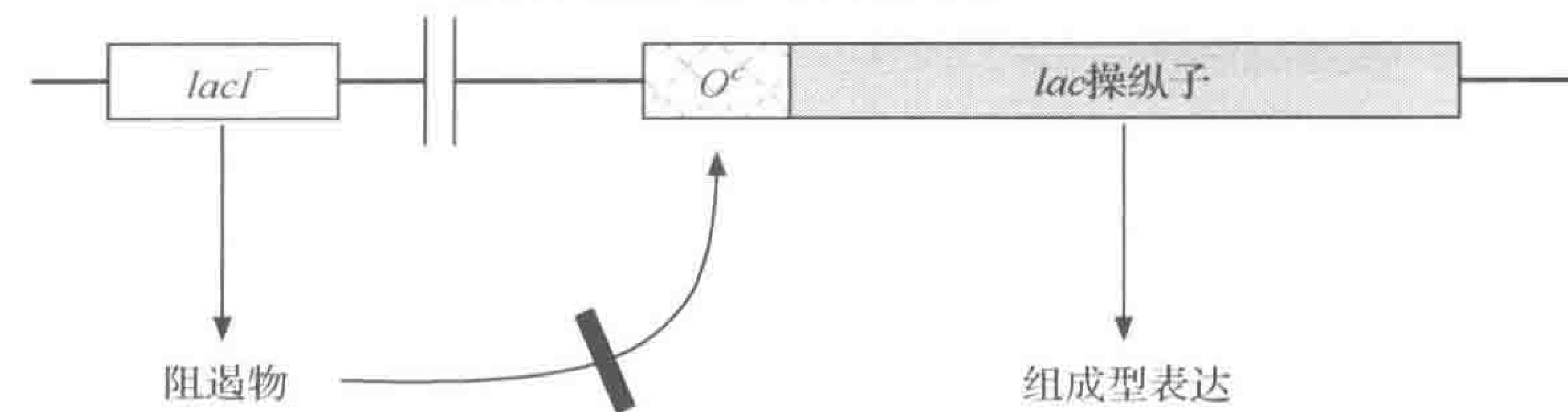


携带功能性*lacI*基因的质粒

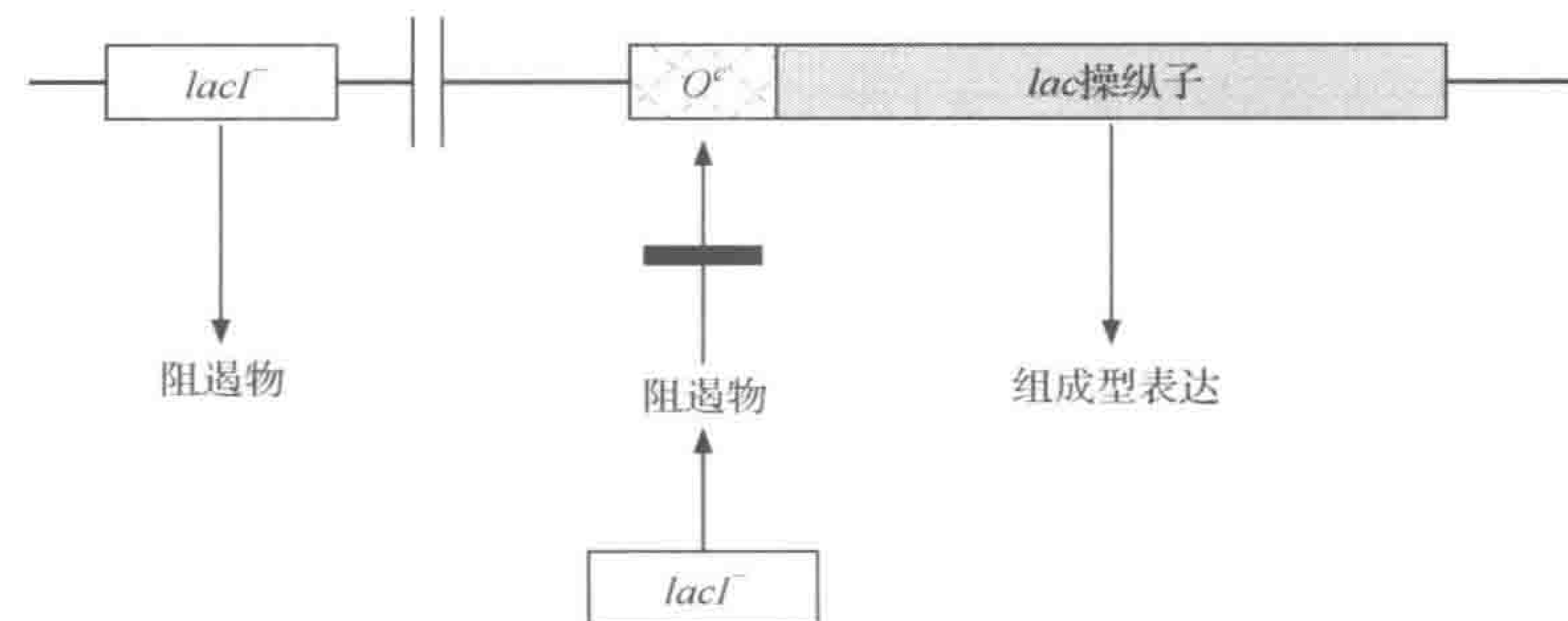
(b) 携带功能性*lacI*基因的质粒恢复诱导性

图 3.13　*lacI* 突变是隐性遗传的

操纵子突变；阻遏物不能结合



(a) *lac*O^c基因突变导致组成型表达



质粒携带功能性*lacI*基因

(b) 携带*lacI*功能基因的质粒不能恢复诱导性

图 3.14　*lac*O^c 突变为顺式显性

正常水平的 *lac* 阻遏物不足以抑制所有多拷贝质粒上的 β-半乳糖苷酶基因时，这些突变在遗传操作中起到重要作用。

代谢物抑制

如果大肠杆菌生长在含有碳、葡萄糖和乳糖作为能量来源的培养基中时，它将优先消耗葡萄糖，直到葡萄糖消耗殆尽时才会利用乳糖进行新陈代谢(图 3.15)。当有葡萄糖存在时，*lacI* 操纵子的基因被抑制。这是一个具有广泛影响的例子，即当一种容易代谢的底物存在时，另一组基因会受到抑制，这就是代谢物抑制。

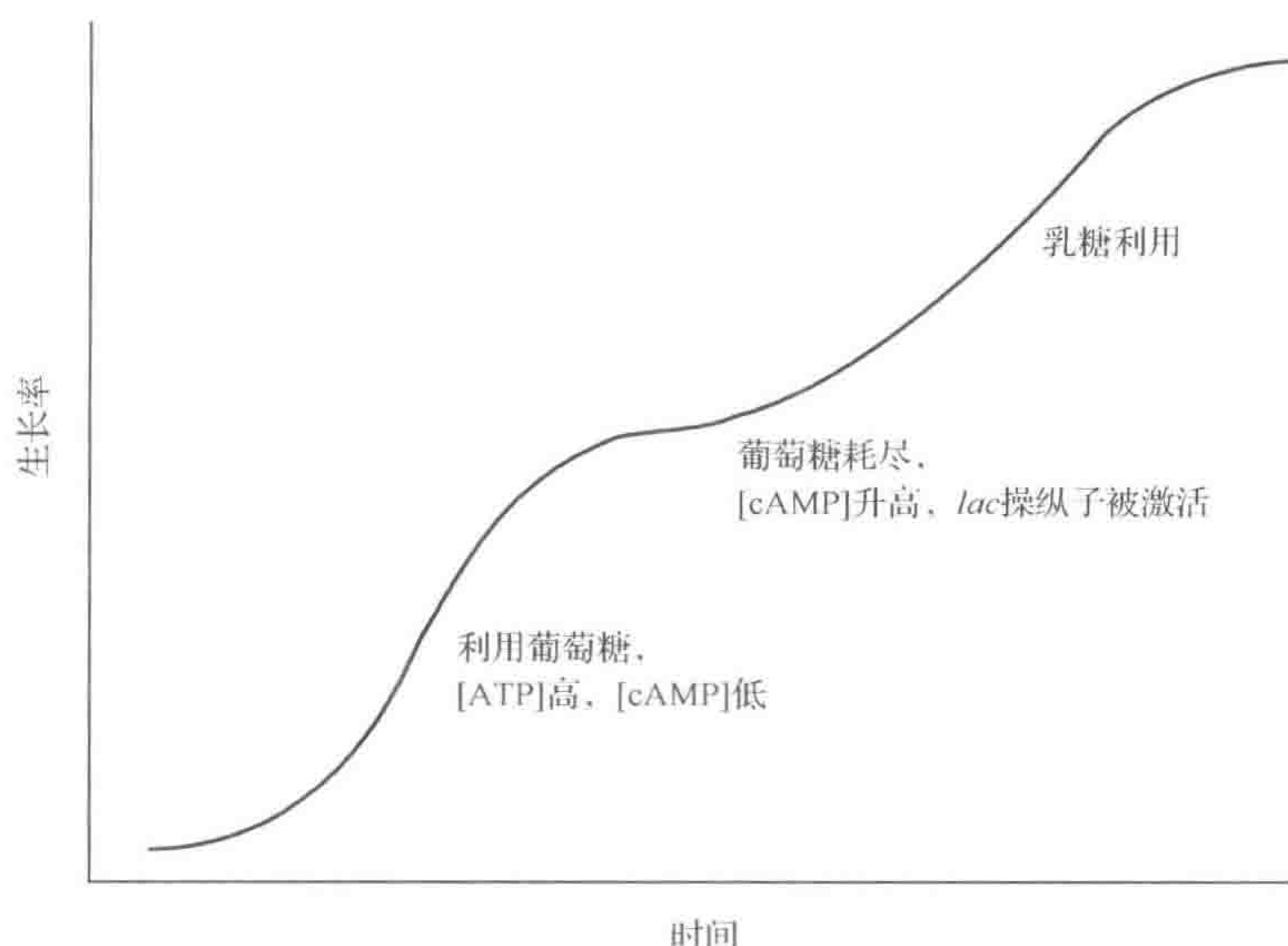


图 3.15 大肠杆菌的两峰生长和代谢物抑制

大肠杆菌 *lacI* 操纵子代谢物抑制的分子基础如下。当葡萄糖存在时，随着葡萄糖消耗释放能量，细胞中的 ATP 水平升高。同时，由于 cAMP 磷酸二酯酶的活化，细胞内的信号分子 cAMP 水平降低。在葡萄糖缺失的情况下，腺苷酸环化酶被激活进而 cAMP 水平提高。*lacI* 操纵子的启动子活性依赖于 cAMP 和 cAMP 受体蛋白(CRP)结合的信号刺激。cAMP 和 CRP 的结合引起了与 DNA 特异性位点识别、结合的蛋白质构象的变化。如果细胞中 cAMP 的水平低，这种刺激和 *lacI* 操纵子的表达都不会发生。

cAMP-CRP 复合物结合到 DNA 启动子上游位点处(−72～−52 是转录起始位点)(图 3.16)，并能够直接与 RNA 聚合酶相互作用，增强后续酶与启动子的结合(图 3.17)。在 cAMP-CRP 复合物缺失的情况下，RNA 聚合酶与 *lacI* 启动子的结合非常弱。尽管这个术语称为“代谢物抑制”，但是应该明确的是 CRP 是正向的，当结合 cAMP 时，它是作为一个激活剂，而不是抑制剂。因此，有人更愿意称它为代谢物激活剂蛋白(CAP)。

cAMP-CRP 复合物结合的另一个作用是使 DNA 在这一点变弯曲(图 3.17)。多数调节蛋白使 DNA 弯曲，这样在螺旋线圈处具有局部影响，并且使启动子更容易与 RNA 聚合酶靠近。

虽然在多种操纵子中如结合位点的位置各不相同等，但其对 cAMP-CRP 复合物的代

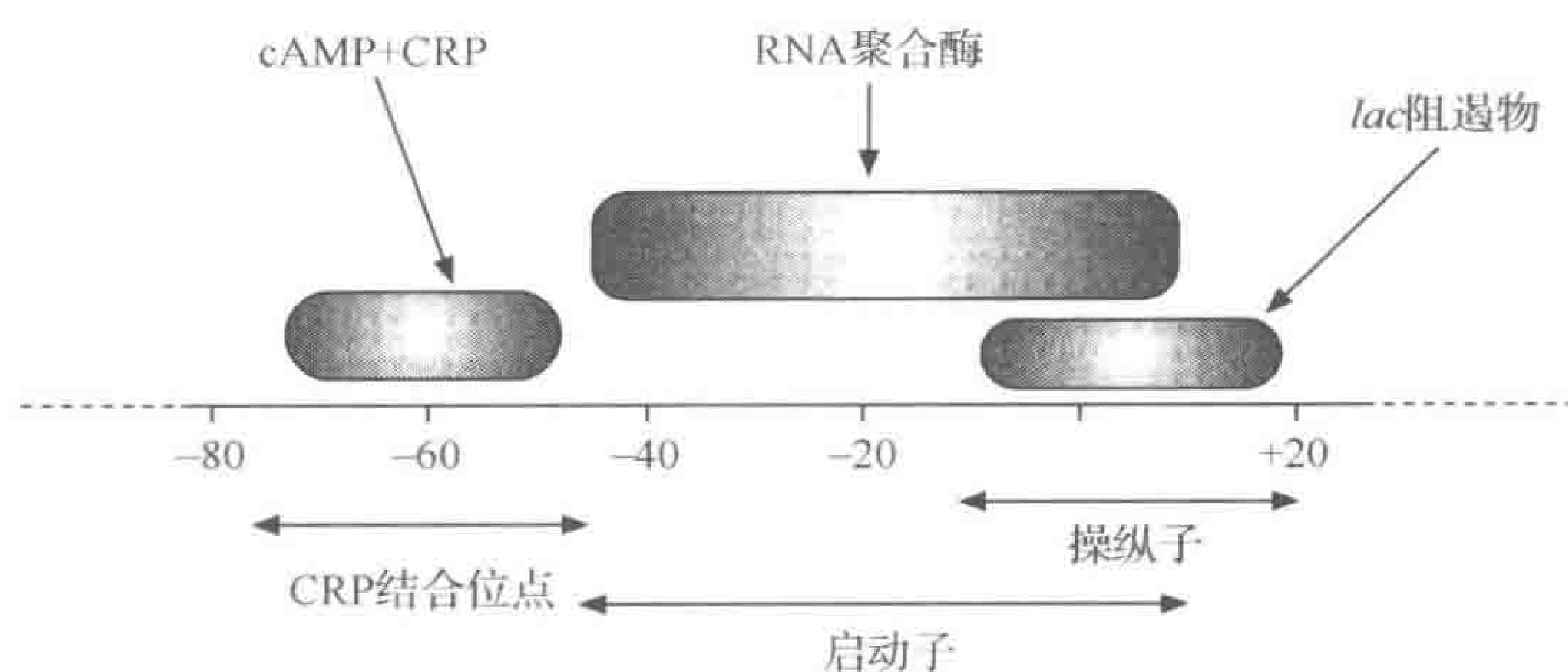


图 3.16　位于乳糖操纵子调节区域的蛋白质结合位点

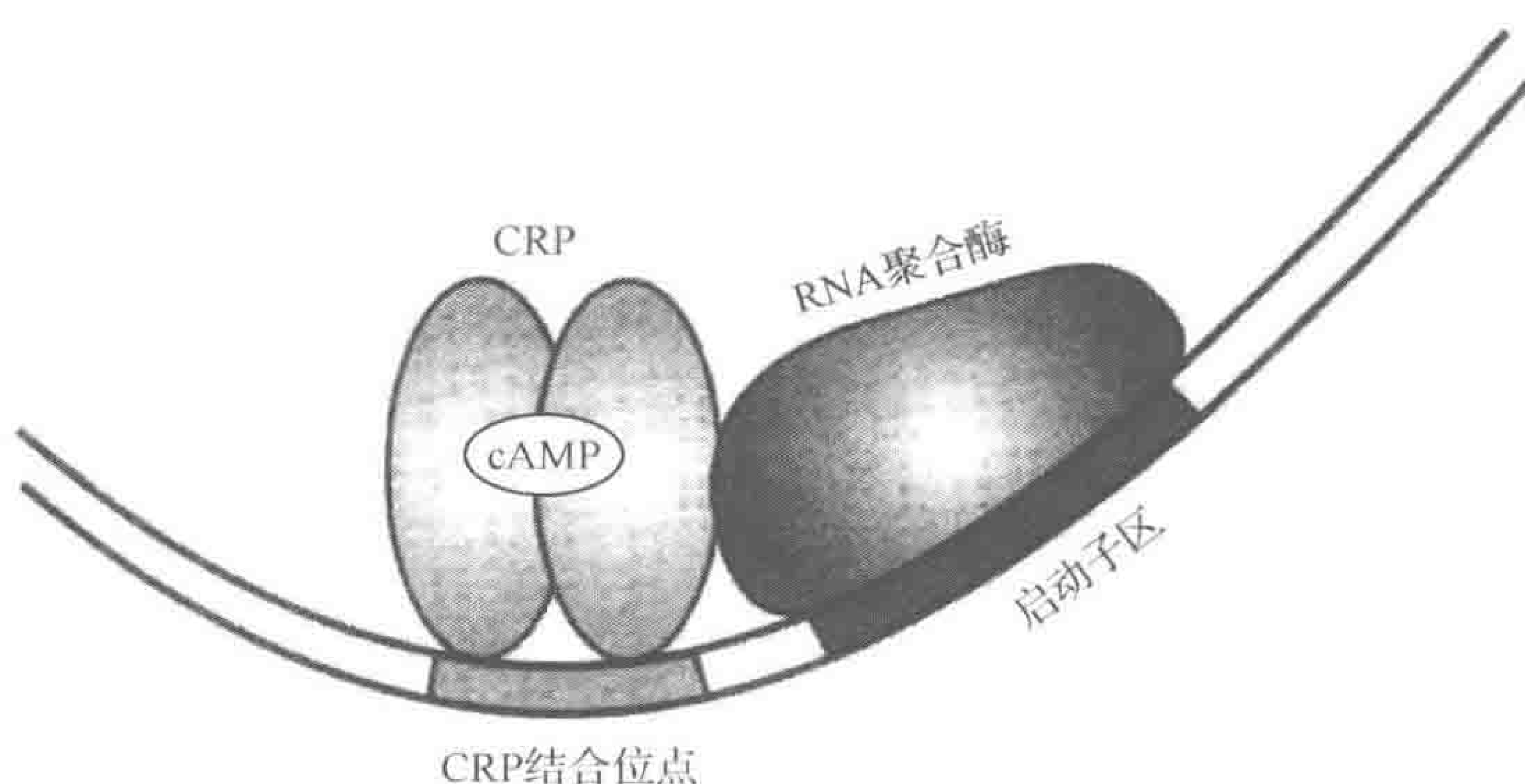


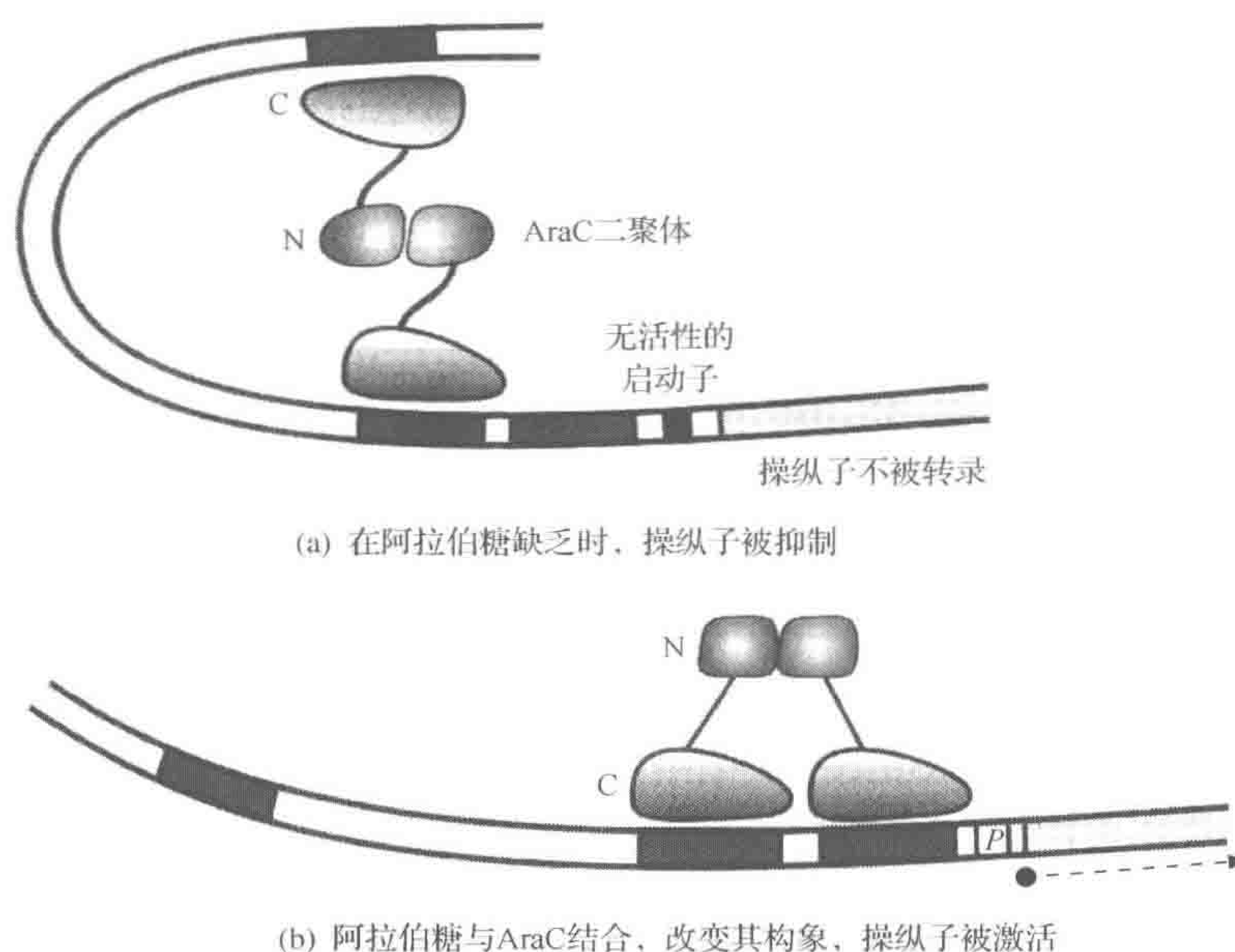
图 3.17　cAMP-CRP 复合物使 DNA 弯曲

谢产物的调节都很敏感。此外，cAMP-CRP 的结合可引起不同的结果，在某些情况下是抑制转录而不是促进转录。

阿拉伯糖操纵子

Lac 阻遏物是一种负调节子，它可以结合到操纵子上阻止转录。然而 CRP(在 cAMP 存在时)是一种正调节子。大肠杆菌阿拉伯糖操纵子的调节蛋白是既可以作为正调节子也可以作为负调节子的例子。

在阿拉伯糖缺乏的条件下，调节蛋白 AraC 的二聚体与 DNA 上分离的两位点结合，形成一个环并且抑制从阿拉伯糖操纵子的启动子处的转录(图 3.18)。环的形成是基因表达调控的重要部分，并且可使远离转录起始位点的相关 DNA 位点参与。AraC 蛋白含有两个独立结构域：N 端结构域结合阿拉伯糖并且介导二聚体的形成，而 C 端结构域含有与 DNA 结合的区域。当阿拉伯糖与 N 端结构域结合时，可改变二聚体形成的方式，进而改变与 DNA 不同位点相接触的能力。结果是 AraC 不再与上游位点结合(DNA 不再形成环)，但是仍与靠近启动子的两邻近位点结合，从而导致启动子的活化。因此，在阿拉伯糖缺乏的情况下，AraC 是一个负调节子，然而在底物存在的条件下，AraC 是正调节子。



(a) 在阿拉伯糖缺乏时，操纵子被抑制

(b) 阿拉伯糖与AraC结合，改变其构象，操纵子被激活

图 3.18 阿拉伯糖操纵子的抑制与激活

trp 操纵子

lac 操纵子和 *ara* 操纵子都是诱导系统的例子，在相关底物存在时开启。另外，当与生物合成通路有关时，如氨基酸的生成，具有相反的调控模式：如果终产物存在时，有关基因的表达应该关闭。这也可以通过调节蛋白的调控来实现，其模型只有很小的改变。例如，负责编码色氨酸合成所需基因的 *trp* 操纵子，在大肠杆菌中由蛋白质 TrpR 调控，TrpR 本身不能与 *trp* 操纵子的启动子区域相互作用，需要色氨酸的结合来形成有活性的阻遏物。

trp 操纵子也被一完全不同的过程即衰减，来调控。因为涉及调控 RNA，将在本章的后续内容中进行讨论。

其他转录调节子

如前面提到的，*lac* 阻遏物属于含有螺旋-转角-螺旋基序的调节子。这是在原核生物中最常见的 DNA 结合蛋白基序。具有多种作用的其他转录调节子也含有这样的基序，这可以用于鉴定基因组序列数据中潜在的调节子基因。在大肠杆菌中，有 314 个转录调节子含有螺旋-转角-螺旋基序。一般来说，这一类的调节子蛋白在螺旋-转角-螺旋基序的指导下与特异的 DNA 序列结合，螺旋-转角-螺旋基序与一调节 DNA 结合活性的更大的调节结构域相连。

在其他例子中，一个被称为 FNR 的蛋白具有与 CRP 相似的结构，能够控制大范围的基因的转录，在缺氧条件下，这些基因的功能为生长所必需。可影响大量基因活性的调节蛋白，如 FNP(和 CRP)，被称为“全局调节子”。FNP 的 DNA 结合活性由氧的可用性调节，即一组铁合硫分子[4Fe-4S]$^{2+}$ 形成真正的氧感受器。OxyR 是这类调节子中的一员，

并且可全局调控过氧化氢压力下的基因表达。在这种蛋白质中，氧化剂的感受器是含有两个半胱氨酸残基的基序。过氧化氢对这些残基的氧化引起分子内二硫键的形成，并且导致构象的改变，这种构象的改变使 OxyR 和其靶基因结合。其他转录调节子的例子在框 3.2 中给出，全局调节子的作用将在本章中的后续内容中详细阐明。

框 3.2 转录调节子

调节子	感受物	活性
LacI	乳糖	乳糖代谢物操纵子的阻遏物
Cap/Crp	葡萄糖的利用	多种代谢物敏感操纵子的激活剂
OxyR	过氧化氢	大肠杆菌中与参与应答 H_2O_2 有关基因的激活剂
PerR	氧化压力	枯草芽孢杆菌中作为参与应答氧化压力有关基因的阻遏物
Fnr	无氧条件	活化参与厌氧电子传递系统有关基因
Lrp	亮氨酸	氨基酸生物合成基因的正调控和氨基酸代谢基因的负调控
Fur	Fe^{2+}	抑制编码铁吸收系统基因的表达
SpxR	超氧化物(通过 SoxS)	活化参与应答过氧化物有关的基因
DtxR	Fe^{2+}	白喉棒状杆菌(*Corynebacterium diptheriae*)中白喉毒力表达的阻遏物，铁吸收调节子
NorR	一氧化氮	抑制大肠杆菌中编码 NO 解毒酶基因
SpvR	进入真核细胞	沙门氏菌属的毒力基因

3.2.4 双组分调节系统

以上描述的系统是对细胞内状态的应答。当没有改变细胞内部状态时，细菌也存在对外界条件和其他刺激进行感受和应答的机制，这种将外部信号传输到细胞内的机制被称为信号转导。

在这些机制中，双组分调节系统最为常见(图 3.19)。一般来说，这些系统含有称为组氨酸蛋白激酶(HPK)的内在膜蛋白和一个称为应答调节子(RR)的分离的细胞质蛋白。HPK 含有两个结构域，输入结构域通常在细胞外面，在理想位置检测环境信号，相反，传递结构域定位于细胞膜的细胞质侧，与 RR 相互作用。当一个刺激物引起 HPK 构象变化时，HPK 在一保守的组氨酸残基处自我磷酸化，并且随后将这一磷酸化基团转移给应答调节子。在此形式中，RR 可与 DNA 结合来调节靶基因的转录。

RR 也含有两个结构域：一个是含有天冬氨酸残基的接收结构域，它可以接收磷酸化基团；另一个是输出结构域，可以与 DNA 结合。应答调节子的去磷酸化对信号的终止是非常重要的，并可以由 RR 本身或一特异的磷酸酶完成。以上的描述对这一类型的感受器进行了传统的概括。然而，因为双组分蛋白的结构域是模块化的，这种方式具有高度适应性，在这种简单的方式中，单一的应答通路含有多个 RR 和其他变异[图 3.19(b)]。虽

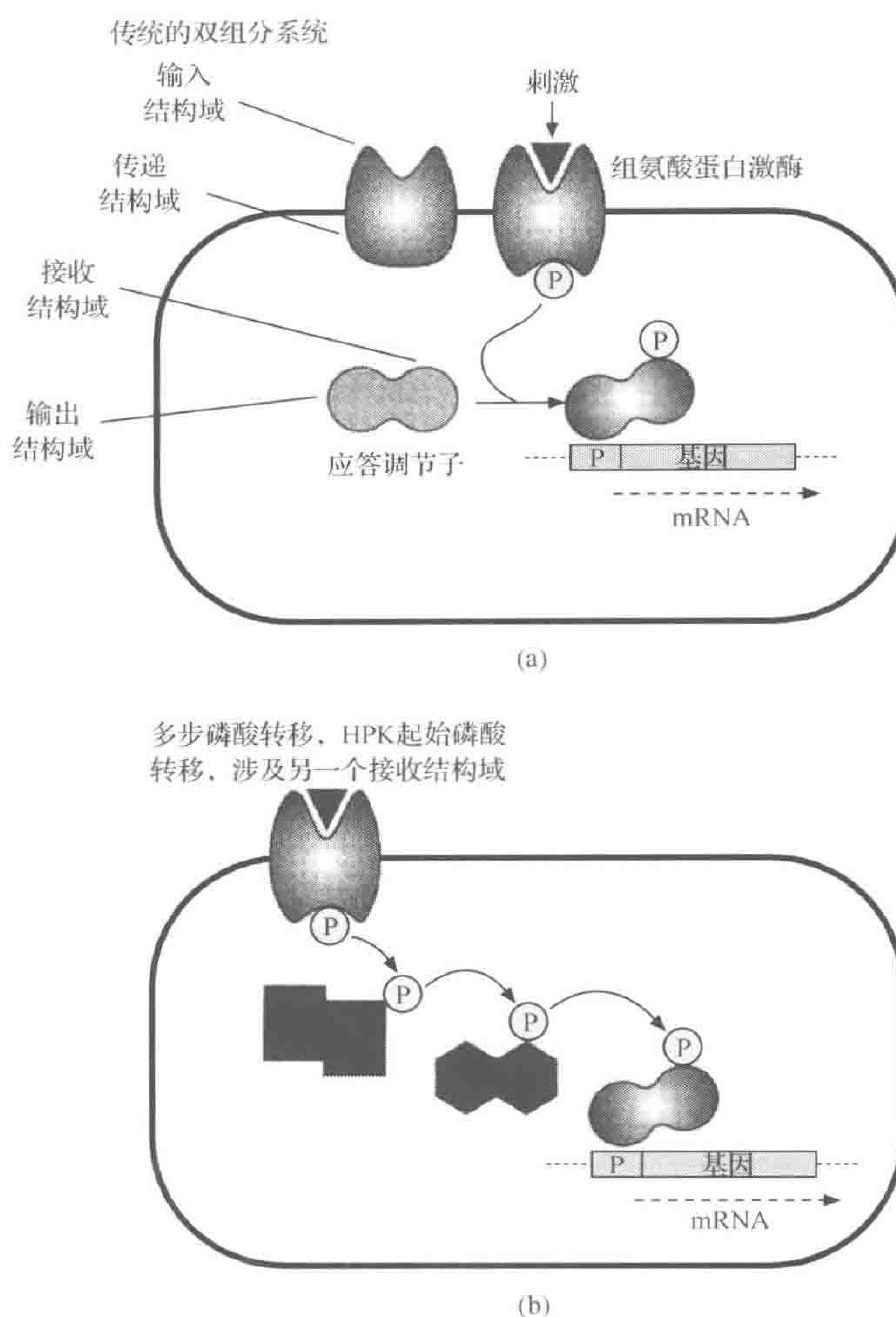


图 3.19　双组分调节系统

(a) 在一传统的系统中，外部刺激物与一特异的膜蛋白相互作用，引起蛋白质构象的改变，这种蛋白质可以激活细胞质侧蛋白质的自我磷酸化，然后这个磷酸盐被传递给应答调节子的接收结构域，使调节子的输出结构域与被调节基因的启动子位点结合；(b) 在磷酸转移系统中，一系列的磷酸化反应激活调节子蛋白 HPK，组氨酸蛋白激酶

然因为历史的原因这些系统一直被称为双组分系统，但是应该注意的是它们也有两个以上组分，并且有时也被称为磷酸化系统。

在细菌中已知的双组分系统已经超过 300 个。例如，大肠杆菌含有 32 个 RR 和 30 个 HPK，铜绿假单胞菌含有 89 个 RR 和 55 个 HPK。只有生殖道支原体(*Mycoplasma genitalium*)不含这些系统。支原体被认为已进化为最小的生命形式，仅含有复制和生长所需的基因，并且为了减小基因组大小已失去了很多调节子功能。

在广泛的双组分系统中，应对不同的刺激，HPK 输出结构域是相似的，但是实际的感应机制的输入结构域具有多样性。例如，FixL[根瘤菌属(*Rhizobium*)得来的 HPK]的输入结构域含有一个血红素基团，它可以应对氧的刺激。然而，植物病原体土壤农杆菌(*Agrobacteriumtumefaciens*)VirA 的输入结构域感受从植物伤口释放的酚类化合物。

在沙门氏菌属(*Salmonella*)中,多效性 PhoP-PhoQ 系统调节毒力和酸的耐受,但是通过周质 HPK、PhoQ 输入结构域真正感受细胞外 Mg^+ 的浓度,它含有多个 Mg^+ 结合酸性氨基酸。其他的双组分调节系统的例子在框 3.3 中给出。

框 3.3 双组分调节系统

生物	组氨酸激酶	应答调节子	感应物	调节
枯草芽孢杆菌	KinA,KinB	SpoOA	?	起始孢子形成
沙门氏菌属	PhoQ	PhoP	Mg^{2+}	表达 Mg^{2+} 转运蛋白和毒力基因
大肠杆菌	EnvZ	OmpR	渗透性	渗透调节外膜蛋白表达
大肠杆菌	ArcB	ArcA	Redox	在无氧条件下抑制有氧酶的表达
大肠杆菌	MarX	NarL	硝酸盐/亚硝酸盐	应答硝酸盐和亚硝酸盐无氧呼吸基因的表达
假单胞菌	TodS	TodT	甲苯	甲苯降解途径
百日咳杆菌	BvgS	BvgA	?	毒力基因表达
土壤农杆菌	VirA	VirG	植物伤口处的化学物质	Ti 质粒基因
大肠杆菌	PhoR	PhoB	磷酸盐	无机磷转移酶

3.2.5 全局调节系统

所描述的调节机制不仅控制单一的操纵子,也包括大量的无关基因。这种系统称为全局调节。例如,前面提到的 cAMP-CRP 系统不只是 *lac* 操纵子的调节子,而且可影响大约 200 个基因的表达。全局调节系统的其他例子包括:热休克反应、酸反应、氧化应激反应、冷休克反应和渗透应激反应。

大肠杆菌的调节网络在所有细胞中是最著名的,包括大约 4400 个基因,其中 300 个基因编码转录因子。仅 CRP、FNR、IHF、FIS、ArcA、NarL 和 Lrp 7 个调节蛋白便负责调节大肠杆菌 51%的基因表达,全局调节蛋白的影响和重要性可见一斑。CRP 是主要的调节蛋白,它通过探测 cAMP 的水平感知能量的状态。与之相似,低密度脂蛋白受体相关蛋白(Lrp)通过感知 L-亮氨酸的浓度来监测细胞的一般营养状况,并相应地调节基因表达。FNR 和 ArcA 分别是调节大肠杆菌对缺氧和氧化还原反应条件的主要调节子。缺氧条件诱导的特定基因进一步由 NarL 调控,它通过对另一亚硝酸盐和硝酸盐电子受体的存在做出反应。

细胞能够应对环境条件变化的另一种方式是通过改变 DNA 超螺旋程度。一般情况下,细胞内超螺旋的整体水平是由引入负超螺旋的 DNA 促旋酶和去除超螺旋的 DNA 拓扑异构酶 I 的活性之间的平衡调控的。然而,DNA 拓扑结构也受类似组蛋白的蛋白质的影响,它可以与拓扑结构结合并引起 DNA 弯曲甚至环绕。整合宿主因子(IHF)和倒位刺激因子(FIS)是全局调节子,它们通过结合 DNA 和调整它的拓扑结构这样的方式来调

节转录。如果把上述信息与一些启动子受 DNA 超螺旋程度的影响结合起来(部分启动子受较高的超螺旋的刺激,而另一部分则更倾向松弛的 DNA),那么很明显,厌氧和周围介质的渗透压的变化等条件将影响全局或局部的超螺旋,进而可能影响相关基因的表达。

很多致病性细菌也使用全局调节系统,以控制它们在宿主生存和逃避免疫反应机制中所需的相关基因的表达。例如,百日咳菌属(*Bordetella*)种能感知不同的刺激,包括温度、镁和烟酸等,并且在应对这些刺激时全局性地调节毒力基因的表达。

这些全局调节系统所涉及的机制是非常多样的,且有时重叠,形成交互调控网络。例如,全局热休克反应是由 σ^{32} 控制。这种互生的 σ 因子调节被称为分子伴侣的蛋白质的表达,后者可以重新折叠受损的蛋白质或降解变性的蛋白质。虽然这些基因通常被称为热休克基因,但它们的表达也受到包括噬菌体感染等其他不同的应激条件的单独刺激。热休克反应也与下文提到的应激反应以及第 2 章所述的由 DNA 易错修复导致的 SOS 反应大幅重叠。

盛宴或饥荒和 RpoS 调节子

直到最近,大多数分子遗传学和代谢的基本概念是源自实验室条件下迅速增长的大肠杆菌细胞。然而,在自然界中,肠道细菌如大肠杆菌生活在可变的营养条件下,因此,进化使能够有效地适应“盛宴和饥荒”的大肠杆菌被保留下来。从对数生长期到平稳生长期的过渡和饥饿,与导致显著抵抗各种应激条件,如热、酸、高盐和氧化应激等的抗性的代谢和形态学的变化有关。调节许多变化和使细胞应对饥荒及不能生长的条件的中枢调节子是另一种 σ 因子 RpoS,它控制 30 个或更多基因的表达。

3.2.6 群体感应

多年来,细菌被认为仅仅是单个细胞,但早在 1905 年,Enwin F. Smith 就写道:“我能想到的唯一的解释是,多数细菌比少数细菌更强壮,因而多数细菌通过联合能够更好地克服对于少数细菌来讲非常大的困难”。这预言性的论述明显地总结了群体感应现象,该现象直到 20 世纪 80 年代才通过细菌计数及其对自身种群密度的反应开始得以阐明。如图 3.20所示,其原理是细菌分泌和响应扩散性的信号。当细胞密度低时,周围介质的信号浓度低,因此,细菌不会做出反应。当细胞密度高时,信号浓度也高,细菌便通过激活一系列特定基因的表达做出反应。

这一现象最初发现于海洋弧菌属(*Vibrio*)中种的生物发光调控研究。这些菌种能够以两个独立的生命周期生活,自由生活的海洋细菌或在鱼和乌贼发光器官占用并提供生物发光的共生细菌。生物发光是高耗能过程,并且在这些细菌中是严格受种群密度控制的。因此,当较多数目的细菌出现时,发光器官发生生物发光,但是当它们以自由孤立的方式生活时,细胞并不能发光。在这些菌种中,被称为自体诱导因子的信号分子是酰基高丝氨酸内酯,它是由酰基高丝氨酸内酯合成酶 LuxⅠ合成并自由穿过细胞膜扩散到周围的环境中(图 3.21)。如果少数细胞存在,那么产生这种化合物是有限的,因此它不能作为信号被识别。然而,当存在同一菌种的许多细胞时(因为更多的细胞能产生它),环境中酰基高丝氨酸内酯的量增加。当它达到一个临界浓度时,它能刺激转录调节子(LuxR),

该调节子能控制生物发光所必需的基因表达。除了海洋弧菌属(*Vibrio*)外,相关的系统已经在超过 50 个革兰氏阴性细菌种类中得到鉴定。

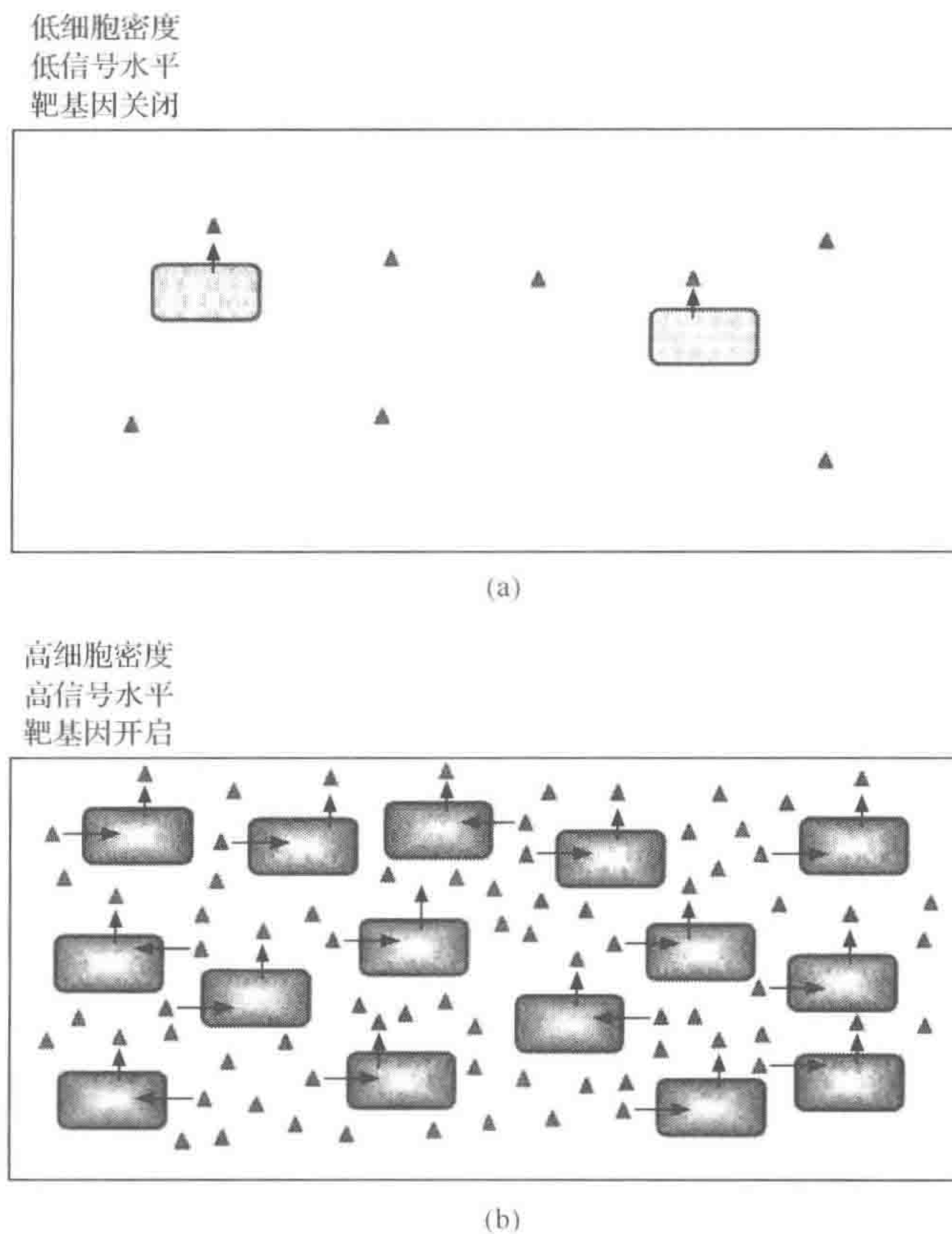


图 3.20 细菌的群体感应现象

细菌某些方面的生理可受其本身浓度的影响。这是受对扩散性的信号浓度的感知而调控的。(a) 当细胞密度低时,周围介质的信号浓度低,细菌不会做出反应。(b) 当细胞密度高时,信号浓度升高。细菌发现这一变化后,相关基因的表达被激活

革兰氏阳性菌在对细胞密度作出反应时也能调节各种过程,它使用小的分泌肽(图 3.22)而不是革兰氏阴性菌使用的酰基高丝氨酸内酯。最初,合成多肽信号前体,后来裂解产生小肽(8～20 个氨基酸)。当这种多肽信号的外部浓度达到临界浓度时,上文提到的双组分系统的 HPK 检测信号并激活刺激靶基因转录的 RR。与革兰氏阴性系统形成鲜明对比的是,这些多肽不能扩散通过细胞膜,它们是特异性分泌的,细胞需要通过信号转导机制对细胞外浓度做出反应。

群体感应促进细菌间多细胞的合作。例如,黏液球菌(*Myxococcus xanthus*)等是捕食性细菌,它通过分泌胞外酶裂解其他细菌获得营养物质。单个黏球菌对细胞的攻击将是无效的,由于消化酶和任何释放的营养物质将会迅速被稀释。因此,黏球菌使用一种"群狼战术的作战单位"的狩猎战略,其中群体感应协调合作喂养大群单个细胞群的形成。作者将在第 9 章中介绍黏液球菌(*M. xanthus*)的其他方面的信息。

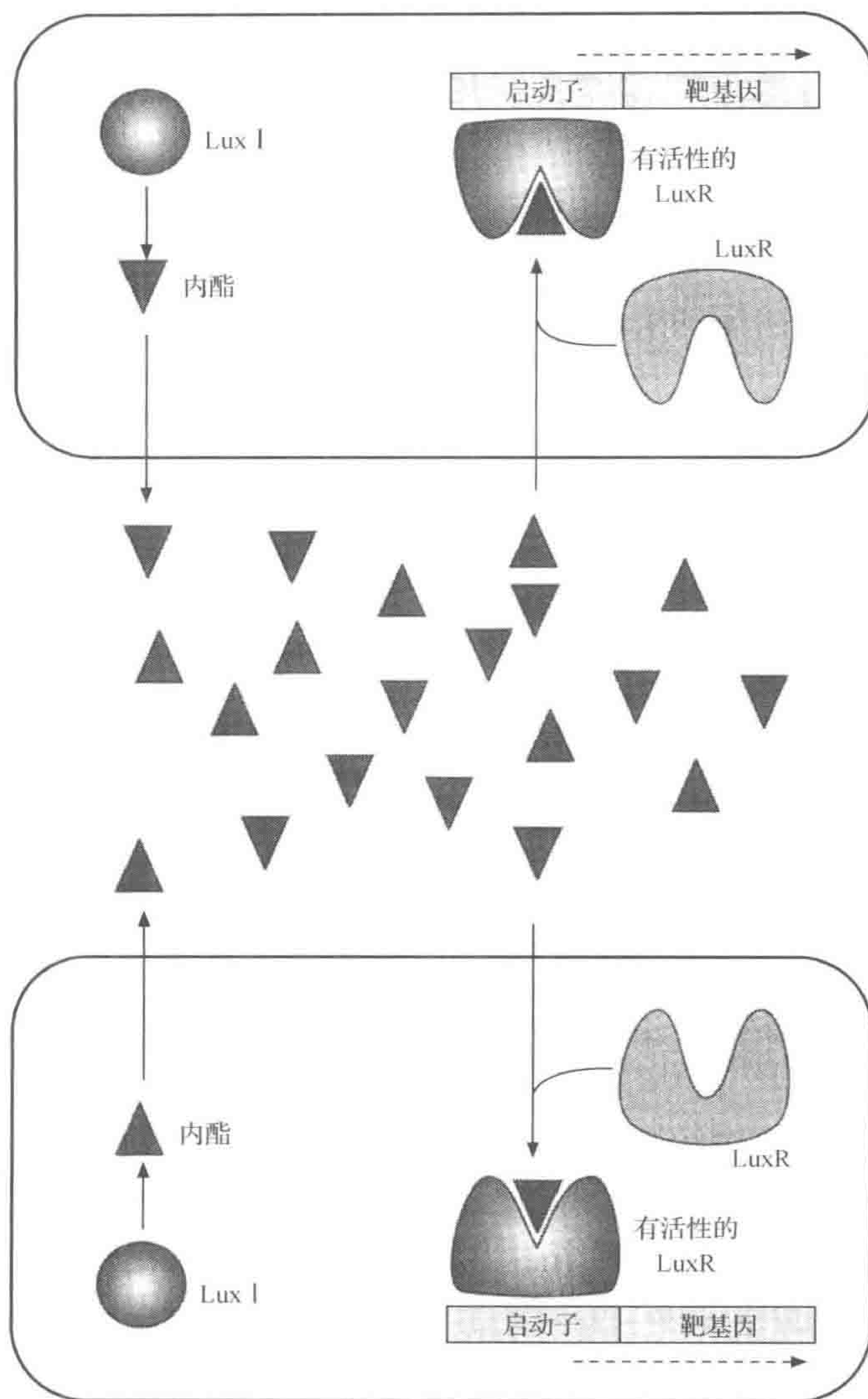


图 3.21 群体感应：LuxⅠ/LuxR 系统

LuxⅠ表达的酰基高丝氨酸内酯介导了许多革兰氏阴性菌的群体感应。内酯能够穿透细胞膜，进入其他细胞并与 LuxR 相互作用。如果内酯的浓度足够高，被激活的 LuxR 将开启相关基因

与其相似，植物病原菌胡萝卜软腐欧文氏菌（*Erwinia carotovora*）使用高丝氨酸内酯群体感应系统，以确保在有足够细胞数量时产生攻击植物结构的消化酶，并对植物组织发起协调一致和有效的攻击。另一个例子，将在第 6 章中看到，细菌间遗传信息转移所涉及的一些机制，同样依赖于群体感应，但仅适用于高细胞密度群体。

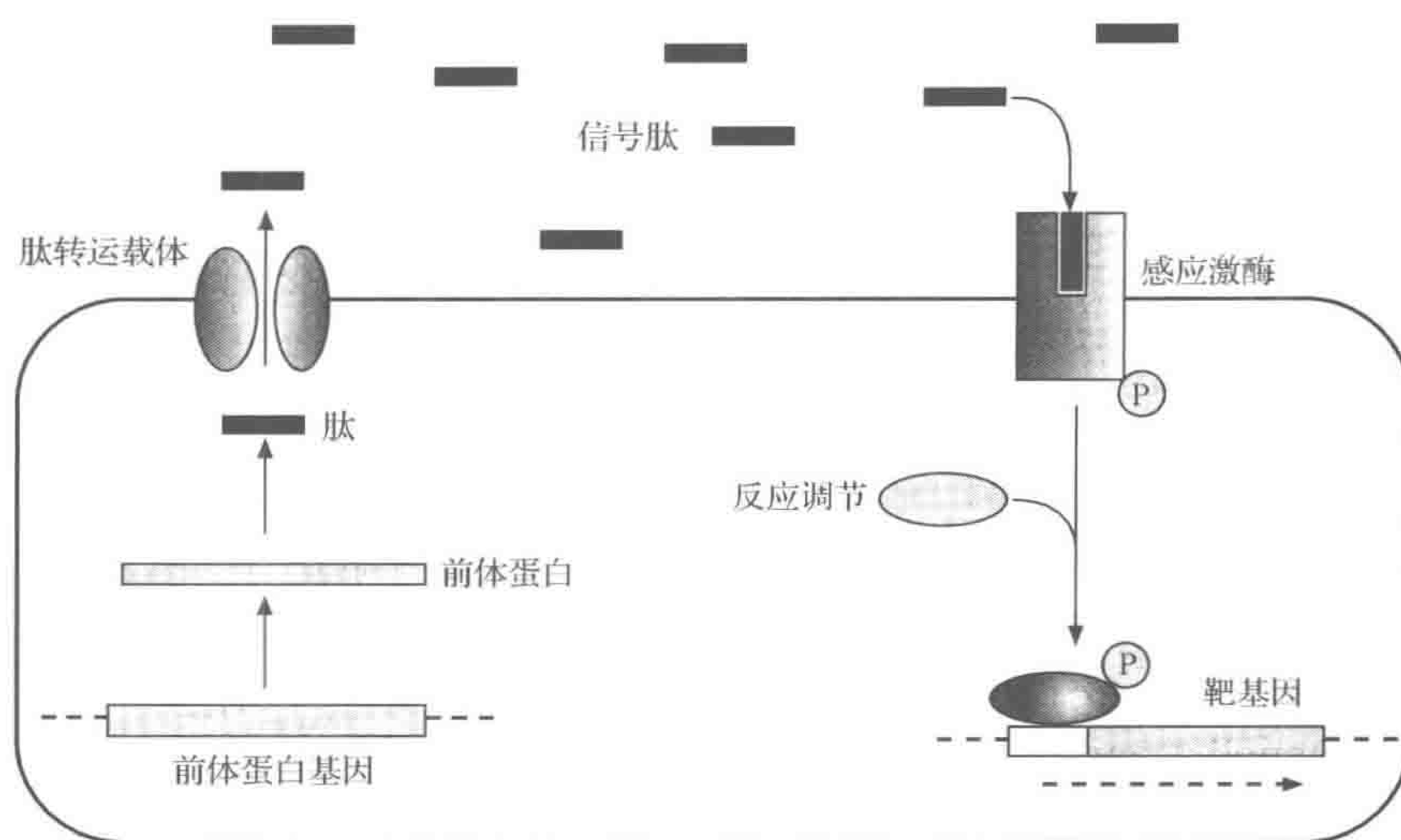


图 3.22 群体感应:革兰氏阳性菌

典型的革兰氏阳性菌的群体感应由特定的肽转运蛋白转运透过细胞膜的小肽介导。靶细胞并不接受这些小肽,但膜受体可以识别这些小肽并将信号传递到细胞内部(也见双组分调节)

3.3 翻译控制

3.3.1 核糖体结合

mRNA 产生后,下一步就是 mRNA 与核糖体的结合,并翻译成蛋白质。这一阶段在细菌自然控制基因表达中似乎发挥一个相对次要的作用,这并不令人惊讶,因为它非常浪费地产生非翻译所必需的大量 mRNA。然而,当特定基因的转录已经达到非常高的水平时,基因工程菌的翻译控制变得非常重要。

mRNA 与核糖体的结合发生在特定的核糖体结合位点(RBS;见第 1 章)到翻译起始密码子 AUG 上游 7 个碱基。在细菌中,这个序列决定核糖体结合 mRNA 的位置,从而确定哪个 AUG 密码子用于翻译起始。

核糖体结合位点(RBS)的实际序列和起始密码子的距离在一定程度上有所不同,因此,正如有弱的和强的启动子一样,也存在弱的和强的核糖体结合位点。然而,RBS 序列并不影响翻译水平:它要么可以与核糖体结合,要么不能与之结合,但从起始密码子到核糖体结合位点的间隔距离,对基因表达具有强有力的影响。

在多顺反子操纵子中,第一个基因翻译后,核糖体可能脱离 mRNA。然后接下来的基因必须有核糖体可以附着的一个位点,以便它能进行翻译。效率有所不同,可能会导致随后基因的翻译效率有所减弱。这是已知的极性效应,是一种替代的衰减机制,即更早地调控操纵子内各个基因的表达水平不同。

如果有一个无义突变导致翻译过早终止,则极性更为显著。这种突变可能会阻止随后的顺反子一起翻译。伸展的 RNA 可以形成茎-环结构,使正在翻译的核糖体缺失,导致转录的过早终止,这些衰减子位点通常会被核糖体的存在屏蔽。

然而在许多情况下，第二个基因的起始密码子和前面基因终止密码子是非常接近的(在数个碱基对内)，甚至可能与它重叠(图 3.23)。在这种情况下，核糖体的 30S 亚基不解离。第一个多肽和 50S 亚基释放后，30S 亚基能接触下一基因的起始密码子重新启动翻译过程。



图 3.23 重叠基因的翻译起始

因阅读框不同，蛋白质 1 的终止密码子(UGA)和蛋白质 2 的终止密码子(AUG)重叠

这一过程的延伸称为核糖体移码，已被证明出现于在没有终止密码子的情况下发生的数量有限的例子中。在这些情况下，当核糖体到达 mRNA 特定位点时(滑动序列，通常包含几个腺嘌呤残基)，它可能会移回一个碱基(a－1 移码)，然后继续多肽合成，但以不同的框阅读 mRNA。这会导致相同的 mRNA 产生两种蛋白质(图 3.24)，其中之一是融合蛋白，它含有两个不同的阅读表部分。由于只是偶尔发生，它代表了实现进一步下调基因表达的一种方式，当只需要极低水平的产物时，这可能是有意义的。仅通过调节 mRNA合成很难获得如此低水平的表达。最典型的例子是来自插入序列流动性的调节(见第 7 章)。

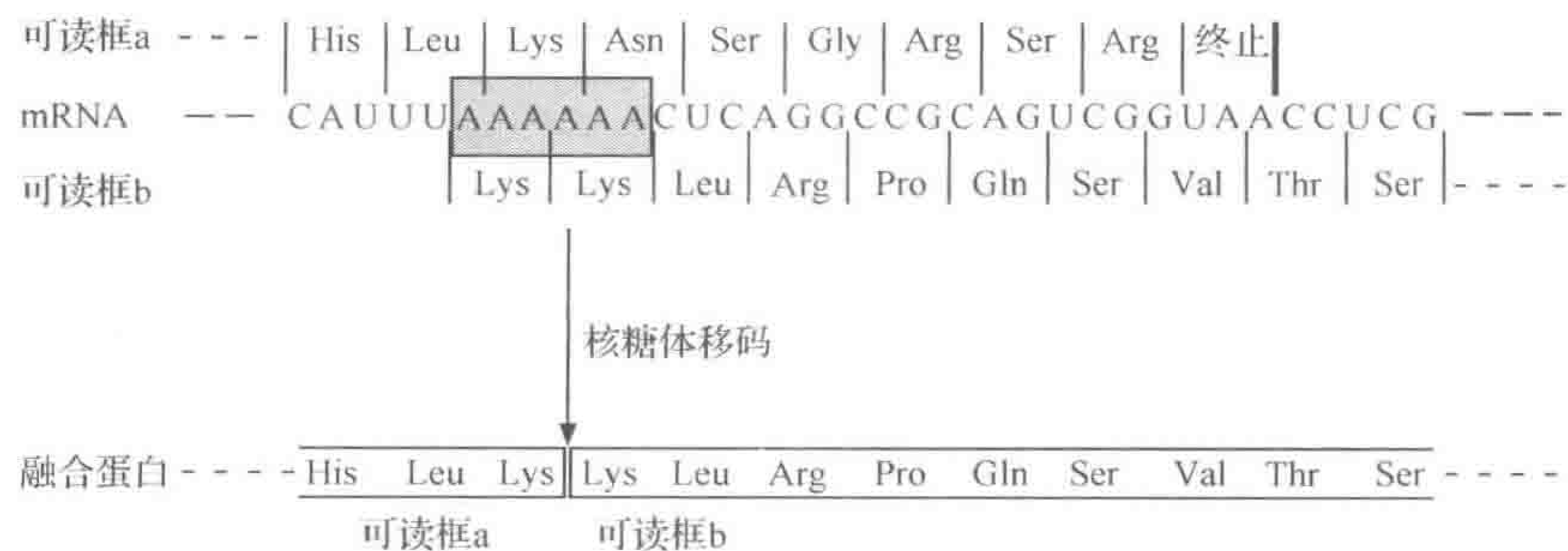


图 3.24 核糖体移码

在箭头指示位置，核糖体可能在 mRNA 上向后移动一个碱基，产生来自两个不同阅读框的一种融合蛋白

3.3.2 密码子用法

整个基因使用的密码子的性质会影响翻译的有效性。大多数氨基酸可由不止一个密码子编码。在某些情况下，因为相同的 tRNA 识别机会相等，密码子是等效的，(见第 1 章)。但在许多情况下，不同的 tRNA 种类负责识别不同的密码子，某些 tRNA 种类以相当低的水平在细胞内存在。包含许多需要这些“稀有”的 tRNA 分子密码子的基因的翻译可能要延误，并最终影响合成产物的量。

这种密码子用法效应的一些间接证据来自对大肠杆菌基因序列的研究，高表达的基因具有较高程度的“密码子偏倚性”，即它们有一个明显的偏爱密码子，能被普通种类的 tRNA 所识别。中度或低水平表达的基因并不能从整体上表现这种密码子偏倚性，它们

可能包含几个相对罕见的 tRNA 密码子。密码子优化似乎对高表达的基因更重要。

3.3.3 应急反应

核糖体蛋白质的合成提供了翻译控制的一个特定例子。这些蛋白质产生的产物需要进行协调,使每一部分等量。从部分上来讲,这是通过自身的控制实现的,即核糖体蛋白的积累将抑制相应的 mRNA 的翻译。这也提供了核糖体蛋白产生和 rRNA 量关联的一种方式。如果 rRNA 短缺,游离核糖体蛋白将开始积累,并关闭相关 mRNA 的进一步翻译。细胞以这种方式应对生长条件的改变。当细胞缺少营养时,它们需要较少的核糖体。这可以通过减少 rRNA 的产物来实现,伴随着 rRNA 的减少,核糖体蛋白的产生随之停止。这是已知的应急反应的一个方面。

由氨基酸缺乏所引起的应急反应,导致无负载的 tRNA 占据了核糖体的 A 位点。已知的应急因子 A 蛋白参与 GTP 向两个不常见的核苷酸[ppGpp(鸟苷 5′-二磷酸-3′-二磷酸)和 pppGpp(鸟苷 5′-三磷酸-3′-二磷酸)]的转换。这些核苷酸阻止 rRNA 操纵子的转录,由于可用 rRNA 量的减少,核糖体蛋白的产生将减少。

3.3.4 调节性 RNA

在细菌细胞中,RNA 有具体明确的功能,如 mRNA、tRNA 和 rRNA。通过引起翻译或转录水平的变化或影响 mRNA 的稳定性等方式,RNA 在调节基因表达方面也发挥着重要作用。这些所谓的调节性 RNA 通过许多不同的机制发挥作用,例如,通过诱导 RNA 构象变化、与蛋白质结合、和其他 RNA 分子的碱基配对或与 DNA 发生相互作用等方式。

几十年前,衰减调控机制被最先发现,但仅在过去 10 年间,调节 RNA 的重要性和普遍性才得到充分的重视。仅大肠杆菌就鉴定了 80 种以上的调节 RNA。这些 RNA 比蛋白质调节子具有潜在的优势,因为它们不需要大蛋白的翻译,因此细胞的生产成本较低,而且也能较快地产生。

衰减作用:色氨酸操纵子

在本章前半部分所提到的 *trp* 操纵子(图 3.25)包括已知的具有 162 个碱基的前导序列,该序列位于转录起始位点和第一个结构基因起始位点之间。前导序列具有一些共同的互补区域,在不同的条件下可形成不同的茎-环结构。在这些不同的二级结构中,只有 3∶4 的茎-环结构能真正地引起终止作用。

前导序列包含一个能翻译成短肽的区域(图 3.26)。当 mRNA 合成达到足够长时,核糖体将结合到 mRNA 上并开始转录。然而,这一序列包含两个色氨酸密码子。因此,当细胞存在色氨酸时才能合成多肽,缺乏色氨酸时,核糖体停留在那一点。当核糖体处在区域 1 的位置,将阻断区域 1 和区域 2 配对,而当随后的区域 3 合成后,区域 2 能和区域 3 自由形成茎-环结构。这一结构(2∶3)不是终止子,并不能阻挡进一步的转录,但它阻止区域 3 和区域 4 配对[图 3.27(b)],进而阻止 3∶4 茎-环结构的形成,若无法形成 3∶4

茎-环结构则转录继续。因此,在色氨酸缺乏时,*trp* 操纵子可以表达。

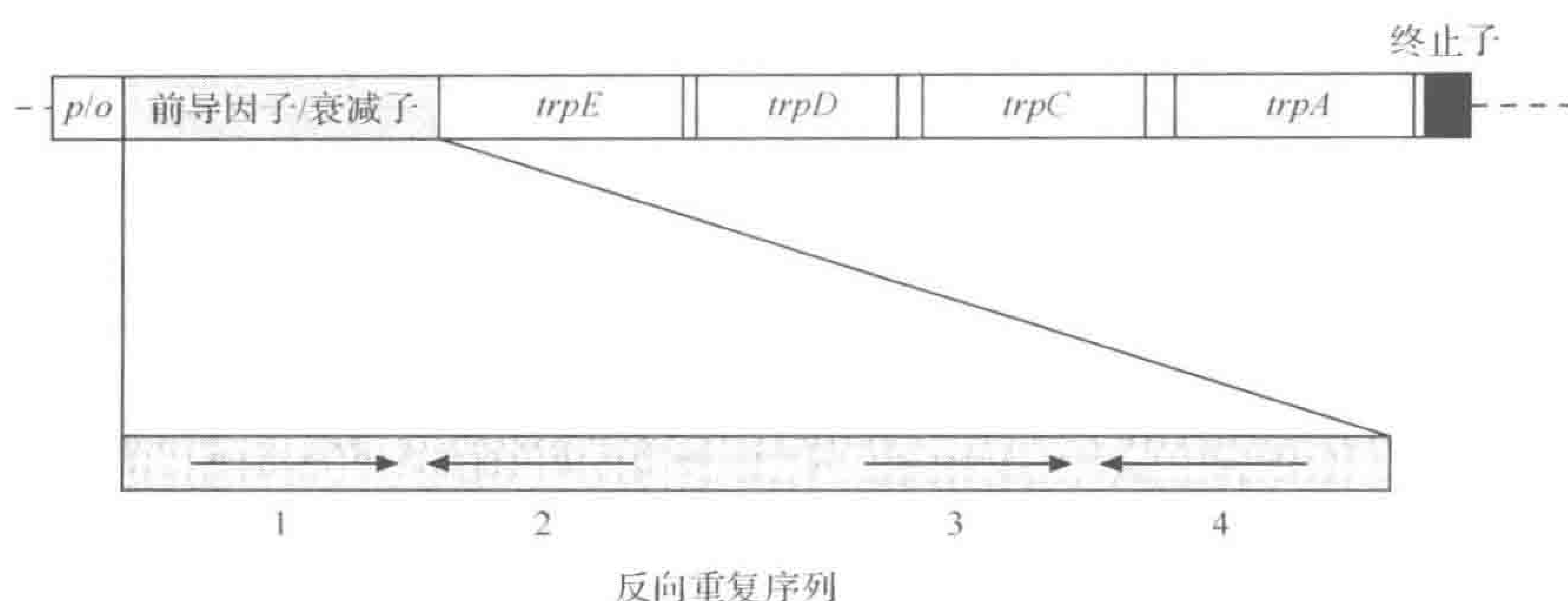


图 3.25 *trp* 操纵子的结构

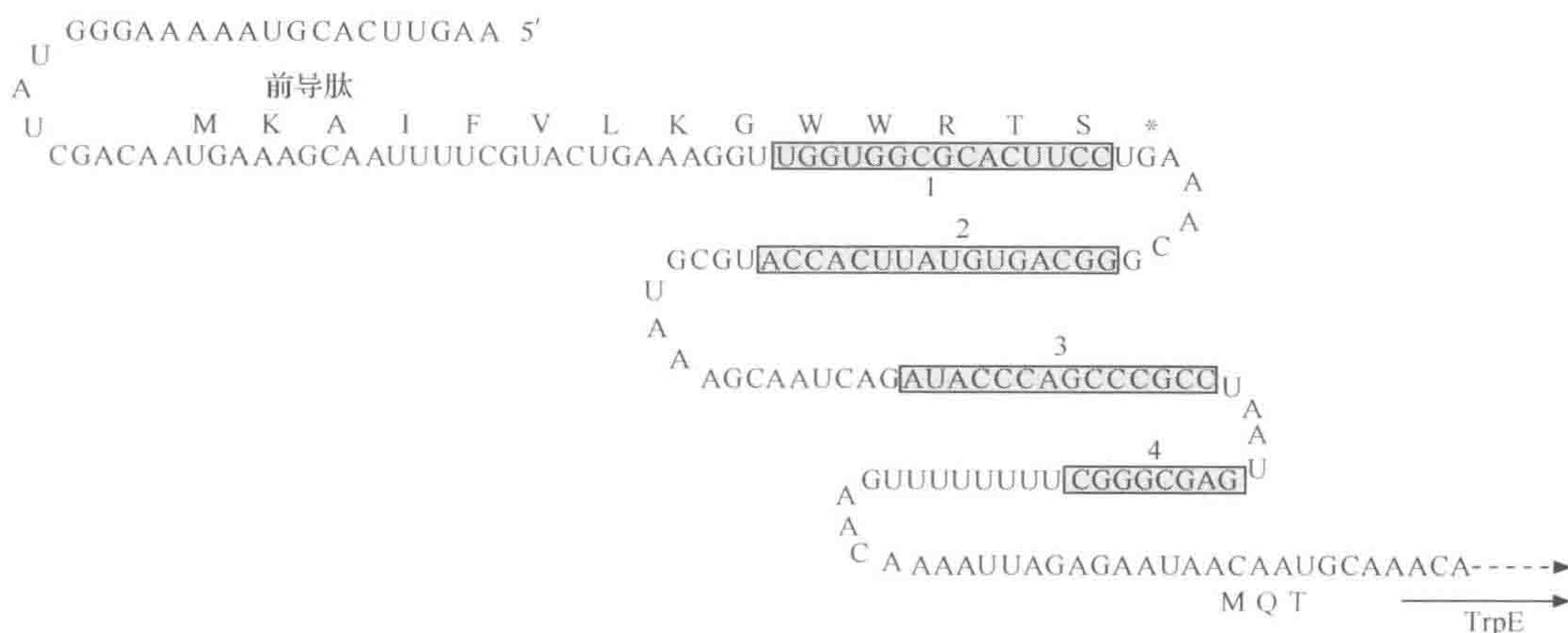


图 3.26 *trp* 操纵子 mRNA 前导-衰减子区域的结构

标记的序列 1～4 是能够形成可替代的茎-环结构的区域

当有足够的色氨酸存在时,核糖体能通过前导序列进行配对直到它们到达终止密码子,和区域 2 的起始部分毗连。虽然核糖体最终从 mRNA 上解离下来,它们占据这一位置足够长的时间以确保阻挡区域 2 形成茎-环结构,如果 2∶3 不能快速形成,区域 4 形成后不久将形成更稳定的 3∶4 结构[图 3.27(c)]。由于 3∶4 结构是鸟苷酸终止子,因此转录终止。当色氨酸存在时将阻止 *trp* 操纵子的表达。相似地,如果蛋白质合成一起被抑制,1 和 2 配对,3 和 4 配对也将导致操纵子的抑制[图 3.27(a)]。

类似的衰减作用方法调控着许多与氨基酸生物合成相关的其他操纵子。在每一种情况下,操纵子含有前导序列,该序列可被翻译成含有相关氨基酸的多种密码子的肽。例如,大肠杆菌的组氨酸操纵子具有前导序列,能翻译成一连串的 7 个组氨酸,而苯丙氨酸衰减子区域具有 7 个苯丙氨酸密码子。核糖体在这些密码子处的减速将阻止终止子结构的形成。

枯草芽孢杆菌的 *trp* 操纵子也受转录水平的衰减作用控制(图 3.28),但是该机制不同于前面所描述的大肠杆菌的调控机制。在这种情况下,茎-环结构出现在色氨酸mRNA

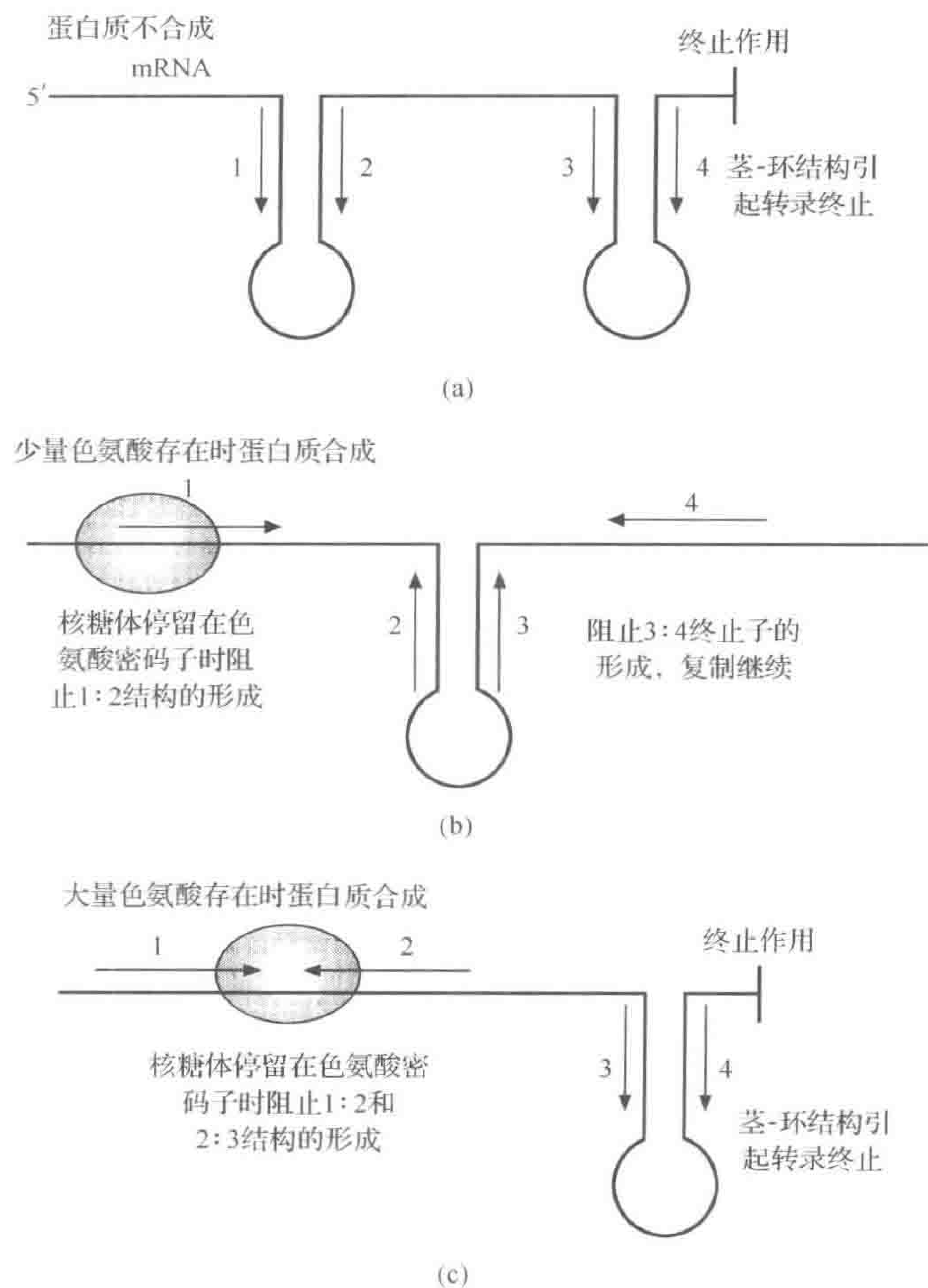


图 3.27　*trp* 操纵子的衰减控制

(a) 在蛋白质合成不存在的情况下，能够形成茎-环结构 3∶4，操纵子不能转录；(b) 如果在数量有限的色氨酸存在的条件下合成蛋白质，核糖体在前导区的色氨酸密码子停顿，阻止茎-环结构 1∶2 的形成。当 RNA 聚合酶转录区域 3，它将和区域 2 配对。结构 2∶3 不是终止子，但它隔离区域 3，从而阻止 3∶4 终止子的形成，并允许操纵子的转录；(c) 在有足够的色氨酸存在的条件下，核糖体将一直转录到终止密码子，阻断区域 1 和 2，这将使 3∶4 终止结构形成，阻止操纵子的转录

的前导序列的 A∶B 或 C∶D 区域，但是仅后者的结构能真正引起终止。在色氨酸缺乏时，由于 B 和 C 之间的重叠，A∶B 结构形成，这将阻止 C∶D 茎-环结构的形成，因此，操纵子开始转录。大肠杆菌和枯草芽孢杆菌系统之间的主要不同点是前导 mRNA 中存在 11 个重复的 DNA 序列(GAG 或 UAG)和被称为色氨酸-RNA-结合衰减蛋白(TRAP)的 RNA 结合蛋白，当存在色氨酸时，TRAP 能够结合这些序列。由于 GAG/UAG 重复序列覆盖整个 A 区域和部分 B 区域，A∶B 茎-环结构不能形成，这将形成 C∶D 茎-环结构终止子。最终，*trp* 操纵子的转录被阻止。

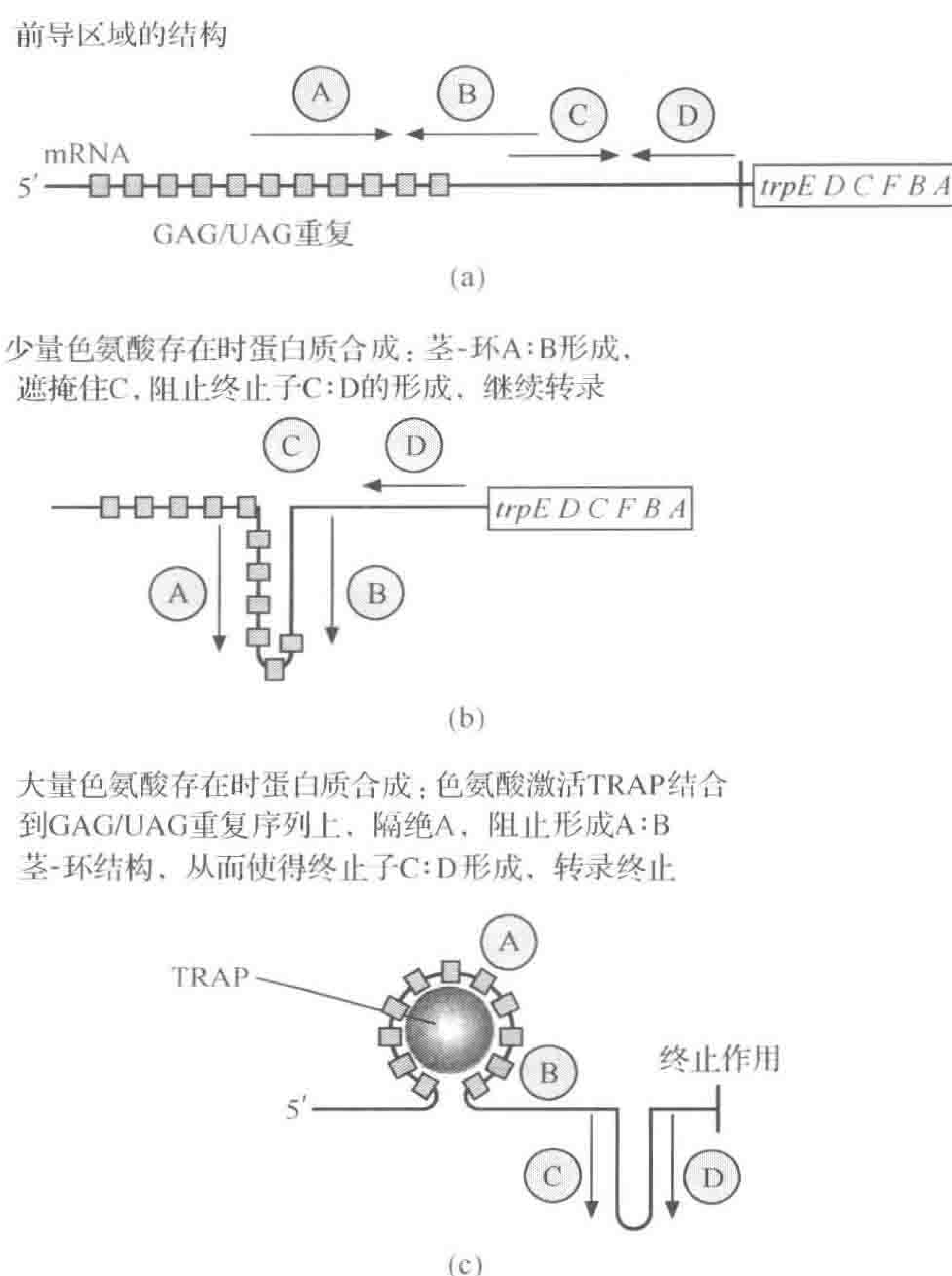


图 3.28　枯草芽孢杆菌 *trp* 操纵子的衰减控制

(a) 前导区含有两对互补序列，使茎-环结构 A∶B 和 C∶D 互相替代成为可能，B 和 C 的部分重叠阻止了两个结构的形成，前导区也含有能够在色氨酸存在条件下结合 TRAP(色氨酸-RNA-结合衰减蛋白)的 11 个重复的 GAG 或 UAG；(b) 在色氨酸存在的条件下，TRAP 不结合，形成茎-环结构 A∶B，不能形成终止子 C∶D，操纵子转录；(c) 在色氨酸存在的条件下，TRAP 结合重复的 GAG/UAG，阻断区域 A 并阻止茎-环结构 A∶B 的形成，区域 C 自由地与区域 D 配对，形成终止结构，阻止操纵子的转录。灰色圆表示被阻止形成茎-环结构的区域

T 盒和 RNA 温度计

在发生衰减作用时，调节性的 RNA 元件位于 mRNA 的 5′端，能对环境信号采取不同构象，这将导致下游基因表达的变化。下面将介绍其他 3 个调节性 RNA 元件，即 T 盒、RNA 温度计和核糖体开关。

RNA 的 T 盒可识别用于蛋白质合成(见第 1 章)的 tRNA 是否被氨酰化，即其是否负载了氨基酸。通常，在编码氨酰 tRNA 合成酶的基因或革兰氏阳性菌氨基酸生物合成相关基因的 5′端非翻译区发现 RNA 的 T 盒。这些元件折叠成能识别特异性 tRNA 的结构，该识别是通过在 RNA T 盒的 3 个碱基和靶 tRNA 的 3 个反密码子碱基之间的碱基配对。基因的表达仅在 tRNA 未被酰基化时激活，因此，T 盒的调节活性促进基因的表达，这些基因为维持用于蛋白质合成的充足的负载 tRNA 池所必需。

RNA 温度计是在 mRNA 5′端非翻译序列形成的另一种类型的调节元件。正如它们的名字所提示的，它们感受温度的变化，并相应地改变基因的表达。在非允许的温度时，RNA 温度计与下游基因的 RBS 结合并折叠成发夹结构，阻止 RBS 被核糖体识别。温度的上升引起结构不稳定，允许核糖体通过，并进行翻译。在大肠杆菌 *rpoH* 基因的上游可以见到这样的元件，它编码 σ 因子 σ^{32}，是热休克反应的重要调节子（见上述的 σ 因子），在这一例子中，调节性 RNA 确保细胞处于破坏性增加的温度时基因表达被激活。

许多细菌性病原菌能够感受到温度的特定变化，从低于 30℃到 37℃，作为已经进入人身体的一种指示，因此，毒力基因相应地得到表达。在致病性的单核细胞增生李斯特菌（*Listeria monocytogenes*）内，RNA 温度计位于编码毒力因子主要控制子上——PrfA 的基因的上游，感受不连续的温度变化，并确保在感染期间细菌生存所需的基因仅在人体温的条件下得到表达。

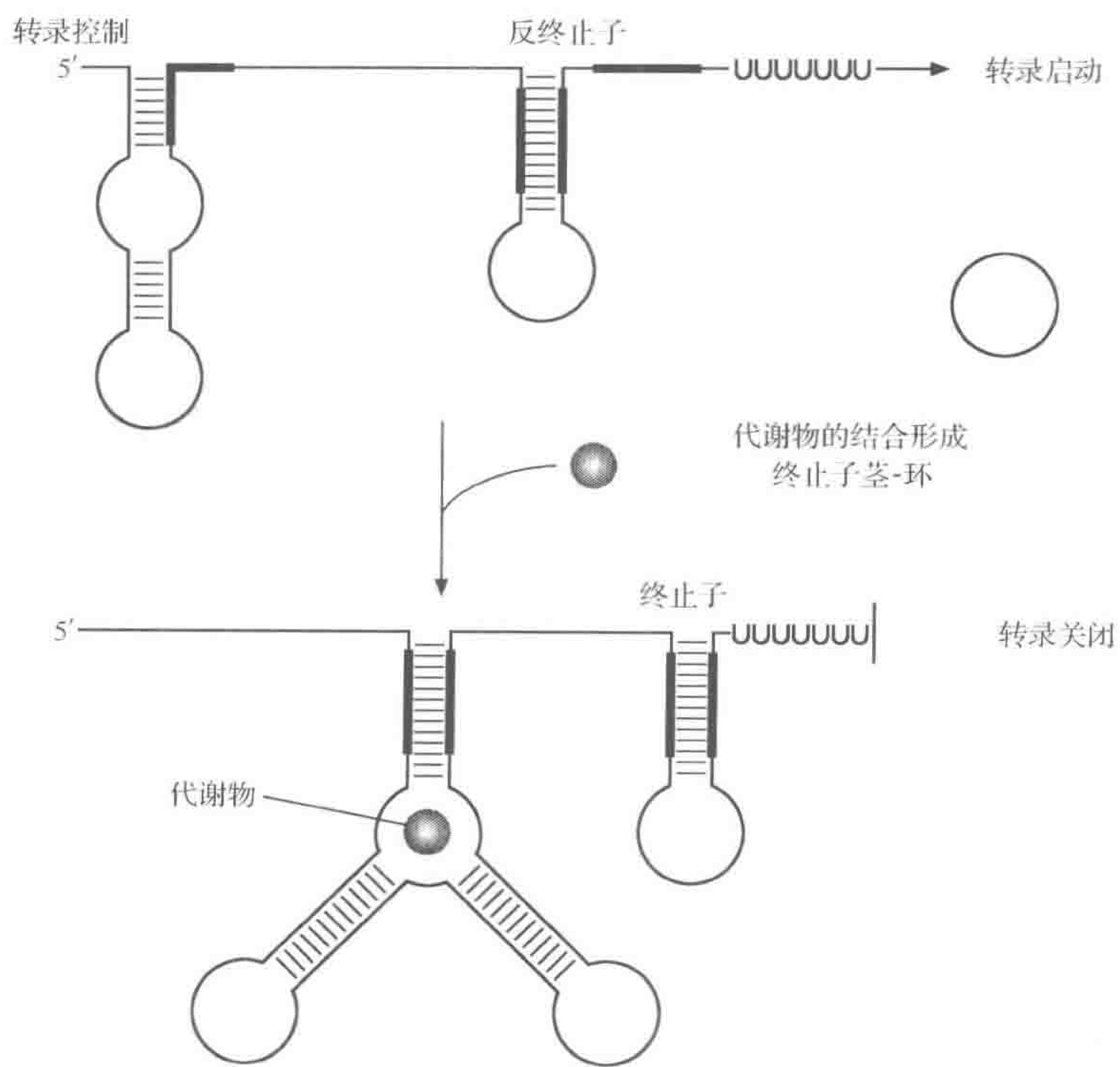
核糖体开关

核糖体开关是基因控制的元件，具有衰减作用机制、RNA 温度计和 T 盒的一些典型特征，鉴别性的特征是它们能结合并感知一定代谢产物的存在。正如 RNA 温度计这一例子，RNA 表现为自身感受器，并不需要如蛋白质因子或 tRNA 的任何中间感受分子。

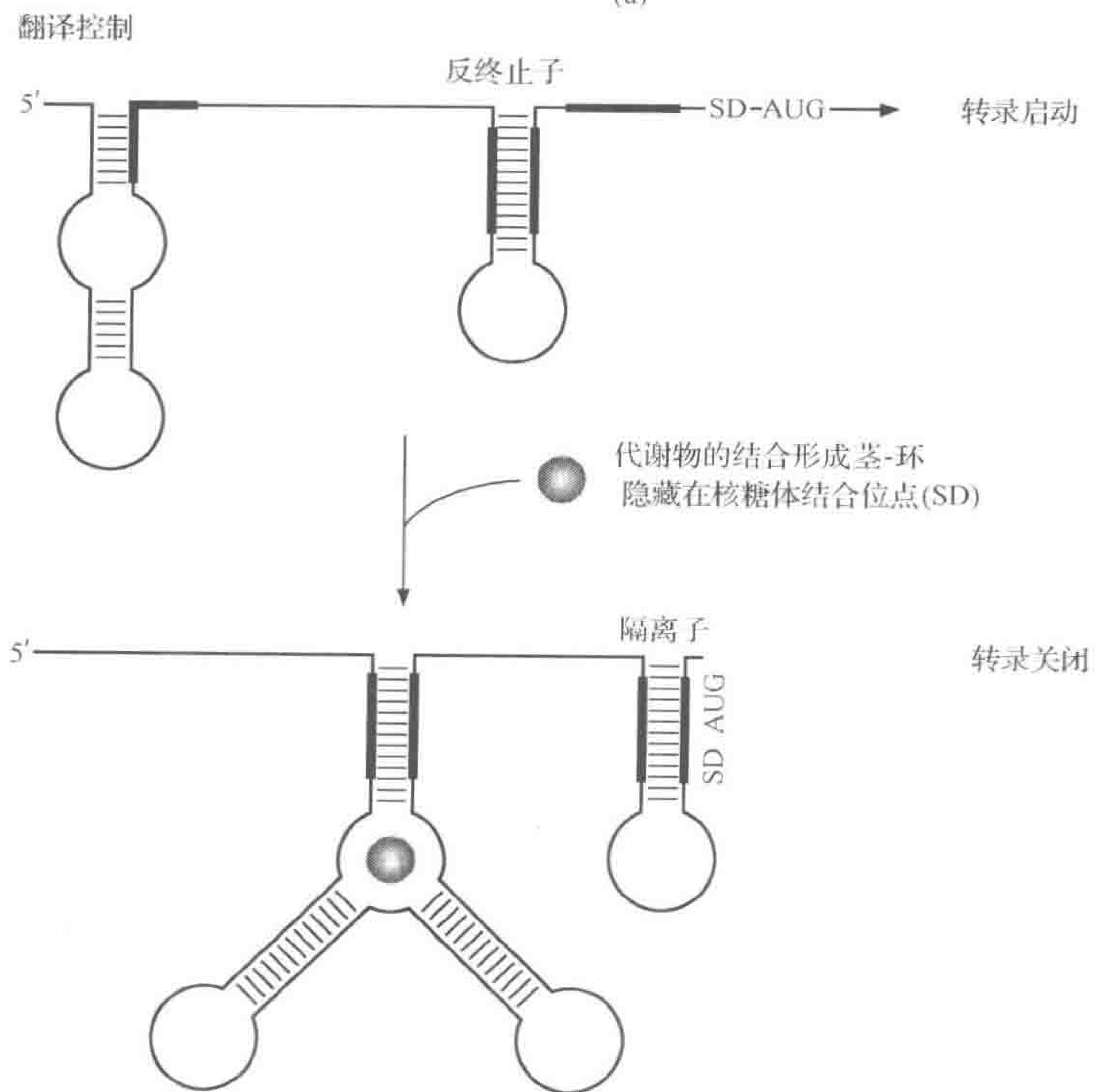
核糖体开关起源于它们所控制表达基因的非翻译调节的前导区域，是由两个相互依赖的并截然不同的结构域组成，一个是负责结合并识别配体的适体结构域，另一个是介导基因表达变化的表达平台。核糖体开关是一种广泛存在和常见的控制基因表达的方法。例如，在具有典型特征的革兰氏阳性菌模型中，2.2％枯草芽孢杆菌可读框具有能转录成核糖体开关的非翻译区，它们调节辅酶类、维生素、氨基酸和嘌呤类的代谢。

当代谢物结合到核糖体开关的适体结构域时，它引起这一区域的 RNA 二级和三级结构的改变，在表达平台内，后者对基因表达的影响的改变发生在转录或者翻译水平。已知的核糖体开关通过 3 种机制起作用。第一种机制[图 3.29(a)]配体结合到适体上，形成转录终止子，并通过诱导终止作用抑制转录。在枯草芽孢杆菌中，具有这种活性的核糖体开关控制维生素硫胺素的代谢。在此硫胺素转换成硫胺素焦磷酸（TPP），它是一种用于糖类代谢的重要酶的辅助因子，TPP 结合到核糖体开关将下调硫胺素生物合成基因的表达以阻止 TPP 的过量产生。

核糖体开关作用的第二种机制是[图 3.29(b)]，配体结合到适体上导致 RNA 复合体的形成，隐藏翻译起始位点并因此阻止下游基因的翻译。核糖体开关作用的最后一种机制[图 3.29(c)]是伴随着配体结合后的核糖体开关 RNA 的自身切割，即另一种称为核酸的 RNA。核酶是一种能催化化学反应的 RNA 分子，可催化核糖体开关自身的磷酸二酯键之一的水解。在枯草芽孢杆菌（*B. subtilis*）中，这样的核糖体开关位于 *glmS* 基因的上游，编码葡萄糖-6-磷酸合成酶。当葡萄糖-6-磷酸含量高时，这种代谢产物将结合到核糖体开关的适体区域并允许表达平台形成核酶，进而导致 RNA 进行快速的自身切割，因此，关闭 *glmS* 基因的表达。



(a)



(b)

图 3.29　核糖体开关介导的基因表达控制

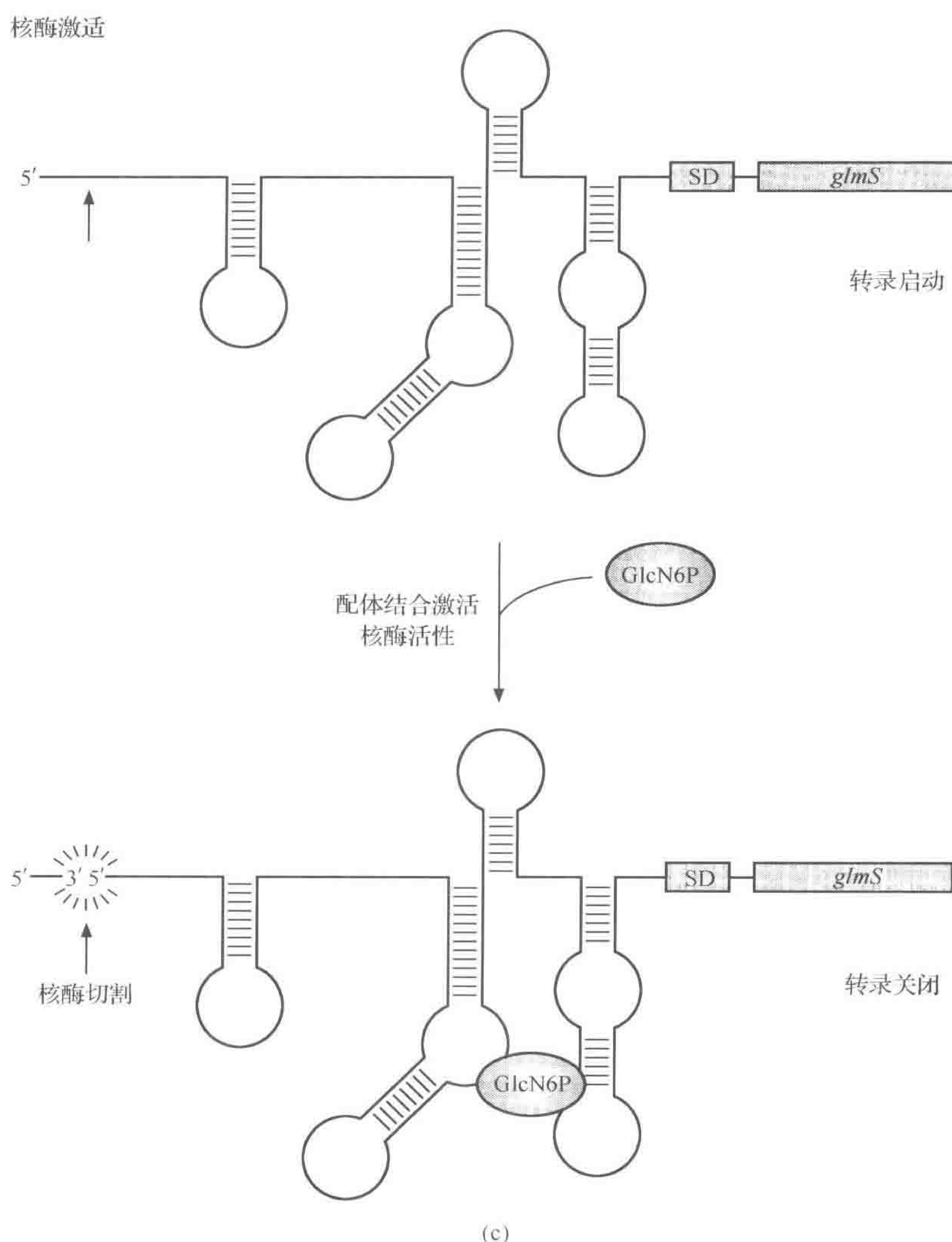


图 3.29 核糖体开关介导的基因表达控制(续图)
(a)配体和核糖体开关的结合引起转录终止环的形成并阻止(b)与(a)相似,但在这种情况下,通过隔离起始位点,在与配体结合后,形成环;(c)配体的结合激活了核酶活性,剪切了 mRNA

小调节性 RNA

在上述提到的例子中,调节性 RNA 元件来自下游基因的 5′端非翻译区,因此它们实际上是调节 mRNA 的一部分。在此,另一组调节性 RNA 分子,是由独立于它们所调节基因之外的基因表达的。这一组包括最大和最广泛研究的 RNA 调节子类型,它们通过和 mRNA 进行碱基配对并通过影响翻译或影响它们的稳定性来调节靶基因的活性。这和 RNA 干扰现象相关,在真核生物中进行了更广泛的研究。

在许多情况下,小 RNA 和它的靶 mRNA 之间的互补受到限制,需要被称为 Hfq 的 RNA 伴侣介导二者间的相互作用,这些分子经常形成部分已知的应急反应系统,通过考

虑一些个别的例子可以更好地理解它们的功能和作用。

RpoS 是另外一种 σ 因子，为了使细胞能耐受压力而调节细胞整体的生理状态（见上述）。RpoS 不仅要对一种压力做出反应，也要对许多不同类型的压力做出反应，小 RNA 分子可实现这些功能。在大肠杆菌中，由于 RpoS 分子 5′端的折叠阻挡了核糖体进入，其 mRNA 翻译处于关闭状态。两个小 RNA 分子 DsrA 和 RprA 能与抑制性结构中的相同区域配对，导致 RpoS 分子 5′端形成的折叠结构不稳定，进而允许核糖体进入并翻译 RpoS。DsrA 的表达是通过低温诱导的，而 RprA 的表达是通过细胞表面的损伤诱导，因此，在两种不同类型的应急状态下，两种小调节性 RNA 使 RpoS 表达，并最终保护了细胞。

小 RNA 也能快速关闭蛋白质的表达并允许细菌细胞在应急状态下重新分配有价值的资源。小 RNA RhyB 具有这样的活性，当必需的铁元素缺乏时，其在大肠杆菌中进行表达。伴随着铁的限制，RhyB 结合并引起编码非必需含铁蛋白的 mRNA 快速降解，通过阻止这些基因的表达，将铁离子重新分配给必需蛋白。

另一个例子是以相对不同的方式进行调控的。CsrA 是大肠杆菌的一种调节性蛋白，在营养缺乏的条件下控制碳源的利用。这种蛋白质能形成二聚体，结合到靶 mRNA 的 5′端非翻译区的特异性 GGA 基序上，因此，这种行为依赖特殊的靶分子影响 mRNA 的稳定性和翻译。CsrA 结合到它的靶序列的能力反过来受两种小 RNA CsrB 和 CsrC 的调节，这也包括 GGA CsrA 结合基序，当这些 RNA 的水平高时，CsrA 偏向结合到这些 RNA 并脱离它所调节的 mRNA。

在第 5 章介绍质粒复制的控制及第 9 章讨论它们在基因操作方面的应用时，将揭示 RNA 调节活性的其他方面。

3.4 相位变异

虽然大多数细菌基因的表达在转录水平上受到控制，且调控幅度可以分级，但是部分基因以“全或无”的方式受到调控。基因表达在开放或关闭状态之间高频的转换称为相位变化，是细菌调节基因表达的另一个重要机制。这将在第 7 章中进行详细讨论。

4 噬菌体遗传学

噬菌体是一种能感染细菌的简单病毒，英文名叫做 bacteriophage（或简称 phage）。噬菌体不仅对分子生物学的发展做出了重要贡献，尤其是在对于基因结构和基因表达的理解中扮演了重要的角色。而且同样重要的是，它们在天然或实验条件下，能使目标基因从一种细菌转移到另一种细菌体内。随着基因克隆的发展，噬菌体还扮演着另一种角色，作为克隆 DNA 的载体（见第 8 章）。

噬菌体和其他病毒的基本特性相似。它们都由蛋白质衣壳及其所包裹的 DNA 或者 RNA 组成。噬菌体结合于细菌表面的特殊受体后，核酸进入细菌体内，从而达到使细菌感染的目的。一些噬菌体基因（早期基因）进入细菌后，几乎立即利用寄主细胞自身的蛋白酶，表达噬菌体 DNA；通常，编码蛋白质需要噬菌体核酸的复制。当一定拷贝的噬菌体核酸产生后，晚期基因开始表达：主要是噬菌体衣壳蛋白合成需要的物质。不同的噬菌体早期及晚期基因表达的转换不同。在以下章节中，将列举一些特殊的例子。然后组装成噬菌体颗粒，菌体裂解，释放大量的噬菌体颗粒，每一个噬菌体颗粒又可以继续感染另一个细菌（图 4.1）。

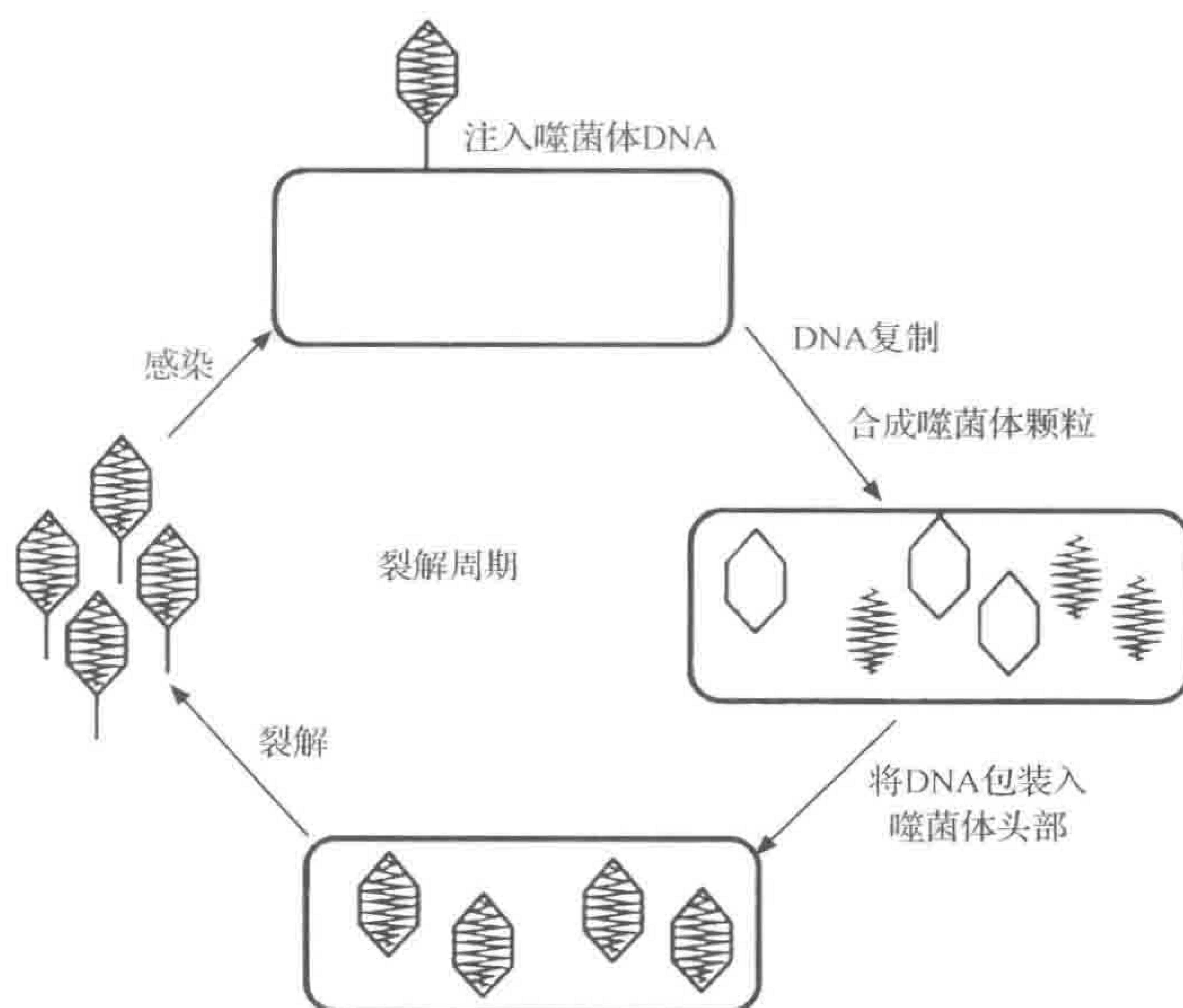


图 4.1　噬菌体的裂解式生长

如果一个噬菌体感染细菌的液体培养物，培养物会由浑浊逐渐变得完全透明。另外，如果细菌在琼脂培养基表面生长，由于每一个噬菌体颗粒首先侵染和裂解一个细菌，然后

以此为中心，再反复侵染和裂解周围大量的细菌，结果就会在菌苔上形成一个具有一定形状、局部透明的噬菌斑。这能够决定噬菌体颗粒在悬液中的密度（噬菌体滴度），即假定样本中噬菌体颗粒的数目与噬菌斑的数目一致。（由于这个假设不总是令人信服的，结果都通常用空斑形成单位表示，即 pfu）。在实际操作中，这类实验通常是将细菌培养液（指示细胞）和噬菌体悬液置于软的顶层琼脂，而不是在平面上扩延。这样就会出现更容易观察的三维噬菌斑。

一些噬菌体，如 λ 噬菌体，通常不进入如上所述的裂解周期，这就是所谓的温和噬菌体。温和噬菌体能与宿主细胞建立共生关系，这种性质通常称为溶源。在这一状态下，几乎所有噬菌体基因都被完全抑制，由于核酸没有被复制，就不会产生噬菌体颗粒，除非溶源状态被再次破坏。在覆盖有适当细菌指示细胞的琼脂上，因为大多数被感染的指示细胞都会被裂解而不是溶源，温和噬菌体通常产生噬菌斑。然而，因为进入溶源状态的溶源菌的继续分裂增殖（能抵抗进一步的感染），噬菌斑会呈现浑浊而不是透明状态，在毒力噬菌体（即噬菌体不能溶源）可以见到这种情况。噬菌体的溶源将会在这一章中得到详细介绍。

4.1 噬菌体的结构

许多噬菌体都按或大或小的范围被分类，本节讨论的所有例子都是能感染大肠杆菌的噬菌体。不同的噬菌体大小和形状都有差异，如图 4.2 所示，个体微小、结构简单的噬菌体，如 φX174 和 MS2；具有可识别的头和尾巴等更复杂结构的噬菌体（λ，T4）；呈丝状结构的噬菌体，如 M13。

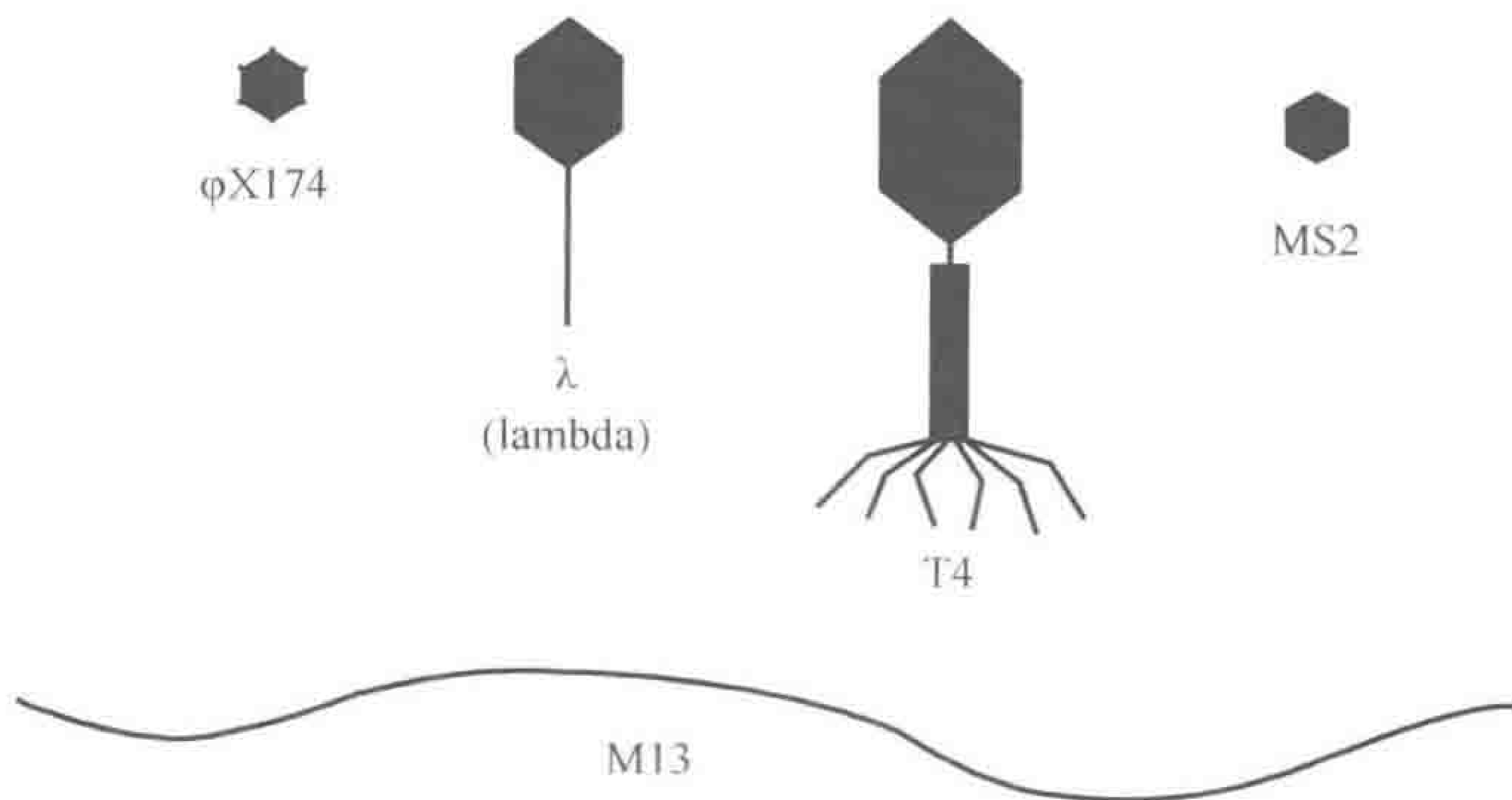


图 4.2 几种噬菌体的形态

请注意 M13 的细丝状结构比在此表示的更长

噬菌体颗粒由蛋白质衣壳和其所包裹的一种核酸分子组成。在结构较简单的噬菌体中，如 MS2，其蛋白质衣壳由大量的一种主要蛋白质自发聚合组成，这种聚合方式决定噬菌体颗粒的大小和形状。从原理上来讲，蛋白质亚单位的自发聚合作用也可以产生丝状结构，但是却不足以决定丝状噬菌体的长度。蛋白质围绕模板聚合确定丝状噬菌体的长度是一种普遍的方式。在接下来的章节将会看到，像 M13 类的丝状噬菌体，噬菌体 DNA 提供了这样的模板。

较大的噬菌体如 λ、T4 的 DNA 被包裹成中空的噬菌体头部，因此，噬菌体头部的大小由组成的蛋白质决定。与之明显相反，M13 的蛋白质围绕着 DNA 聚合形成噬菌体衣壳。这些结构的大小和复杂程度决定了其形成的不同方式，由于自发聚合作用效率太低，直到聚合完成，其结构一般还不稳定，因此，组装噬菌体的头部需要许多额外蛋白质的临时存在，这些蛋白质会在噬菌体最终成熟前被移除（图 4.3）。这种只参与组装而在最终的噬菌体颗粒中并不存在的蛋白质成分叫做支架蛋白。

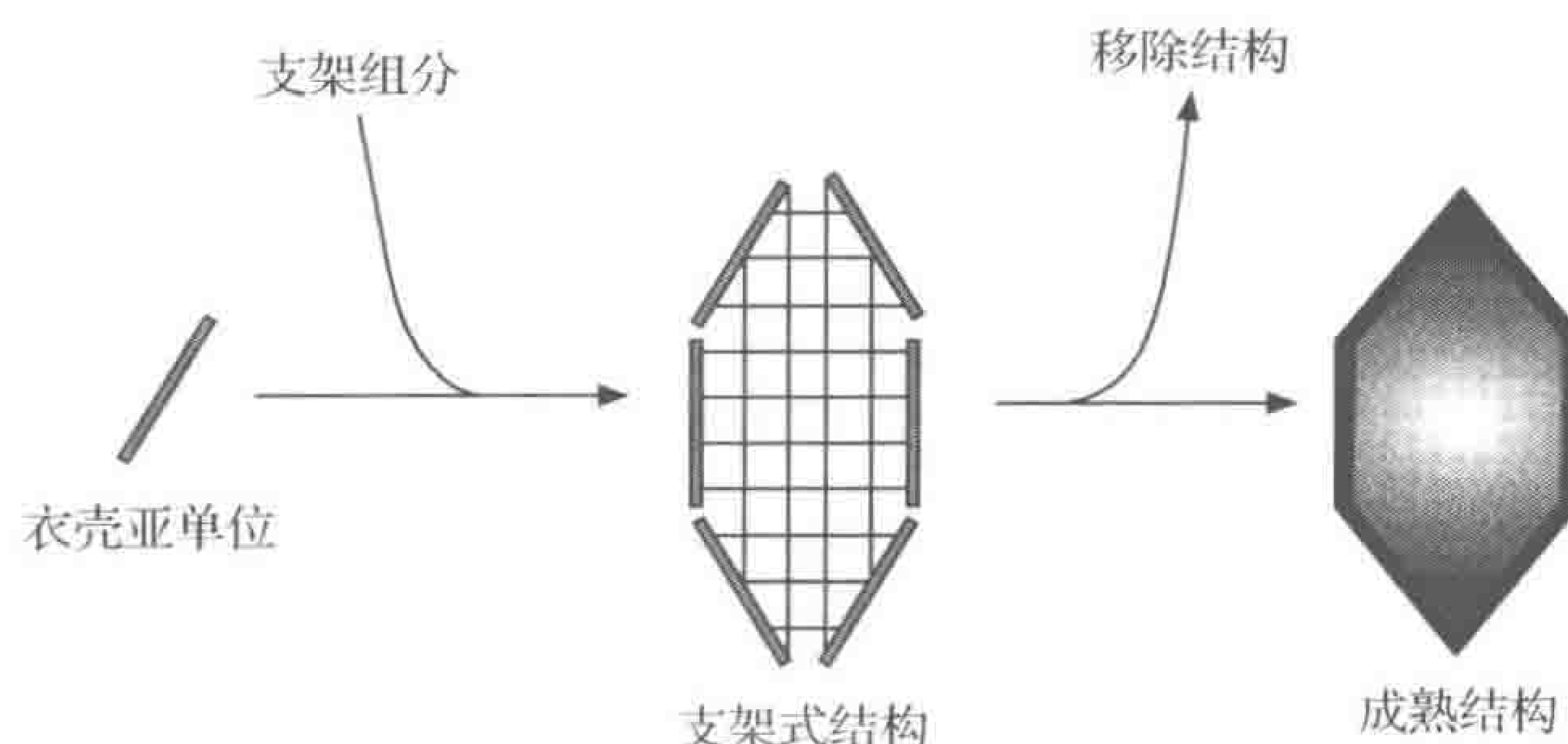


图 4.3 支架式装配

复杂的噬菌体装配需要支架蛋白的参与，一旦装配完成，这些支架蛋白即被移除

在核酸的成分方面，噬菌体也是不同的。φX174 和 M13 含有环状单链 DNA，MS2 含有 RNA 基因组，λ 和 T4 颗粒中含有双链 DNA。由于这些噬菌体提供了开展基础性分子过程研究的例子，每一种噬菌体都值得更深入的研究。

4.2 单链 DNA 噬菌体

4.2.1 φX174

φX174 是一种二十面体的噬菌体，它包含一条由 5386 个碱基组成的环状 DNA 单链，编码 11 种蛋白质，每一种蛋白质都已被鉴定。所有的蛋白质加在一起共由 2330 个氨基酸组成，理论上需要 6990 个碱基编码（3×2330），实际上长于基因组总长。通过基因组序列的检测，很容易发现这种情况是如何形成的（图 4.4）。

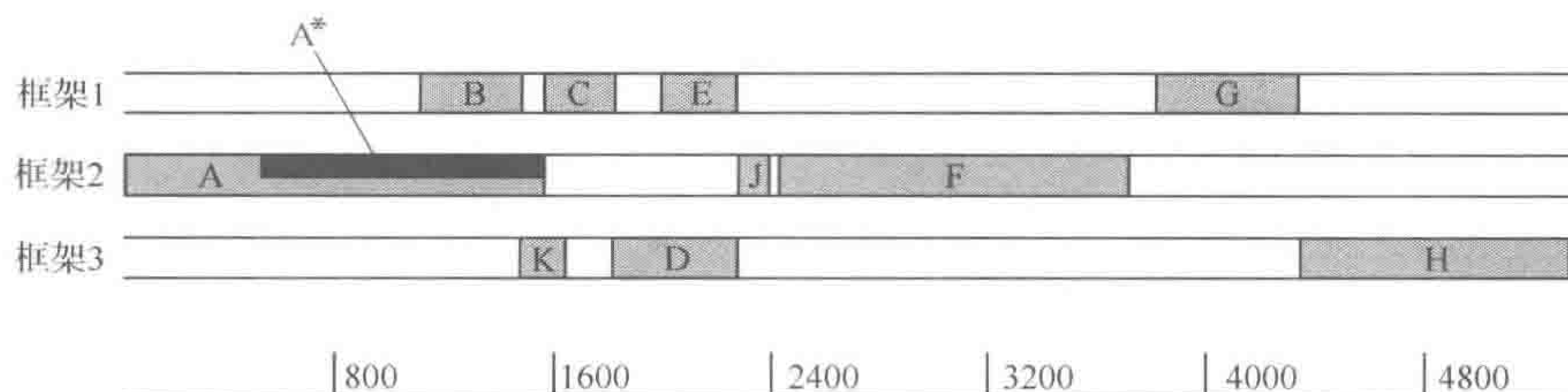


图 4.4 φX174 的基因组结构

实际上 DNA 是环状的，所以请想象其末端是相连的

首先，基因被紧密包裹，整个基因组中非编码序列极少。很多时候，一个基因的末尾紧邻（或者有轻微的重叠）另一个基因的开始。其次，因为其翻译起始位点位于 *A* 基因内

部，其中一种蛋白（A*）对应蛋白 A 的 C 端区域。但是即使忽略蛋白 A*，编码蛋白质所需的碱基仍然超过基因组长度。这是由大量的重叠基因导致的。因为遗传密码按每 3 个碱基编码一个氨基酸阅读，通过可能的阅读框，每一段 DNA 序列理论上均能编码 3 种不同的蛋白质（实际上，双链 DNA 通过双向转录，理论上能同时编码 6 种蛋白质，但是 φX174 基因组只按一个方向转录）。例如，φX174，蛋白 B 由同样编码蛋白 A 的部分碱基编码，它们的编码框不同，所以产生完全不同的氨基酸序列。相似的，基因 *E* 完全在基因 *D* 中，基因 *K* 与基因 *A* 和基因 *C* 重叠，它们的编码框互不相同。遗传密码的有效利用是一些微小的细菌病毒的共同特征，利用了细菌具有由一个多顺反子信号引起的多位点起始翻译的蛋白质合成机制的能力。由于相关基因的进化演变可能补偿这种机制，在个体较大的病毒和细菌基因组中这种重叠基因的使用范围有限。

因为噬菌体颗粒只有一条 DNA 链，与细菌染色体复制相比，噬菌体的复制具有明显的不同，如图 4.5 所示。当 φX174 的单链 DNA 进入靶细胞后，通过合成其互补（或“负”）

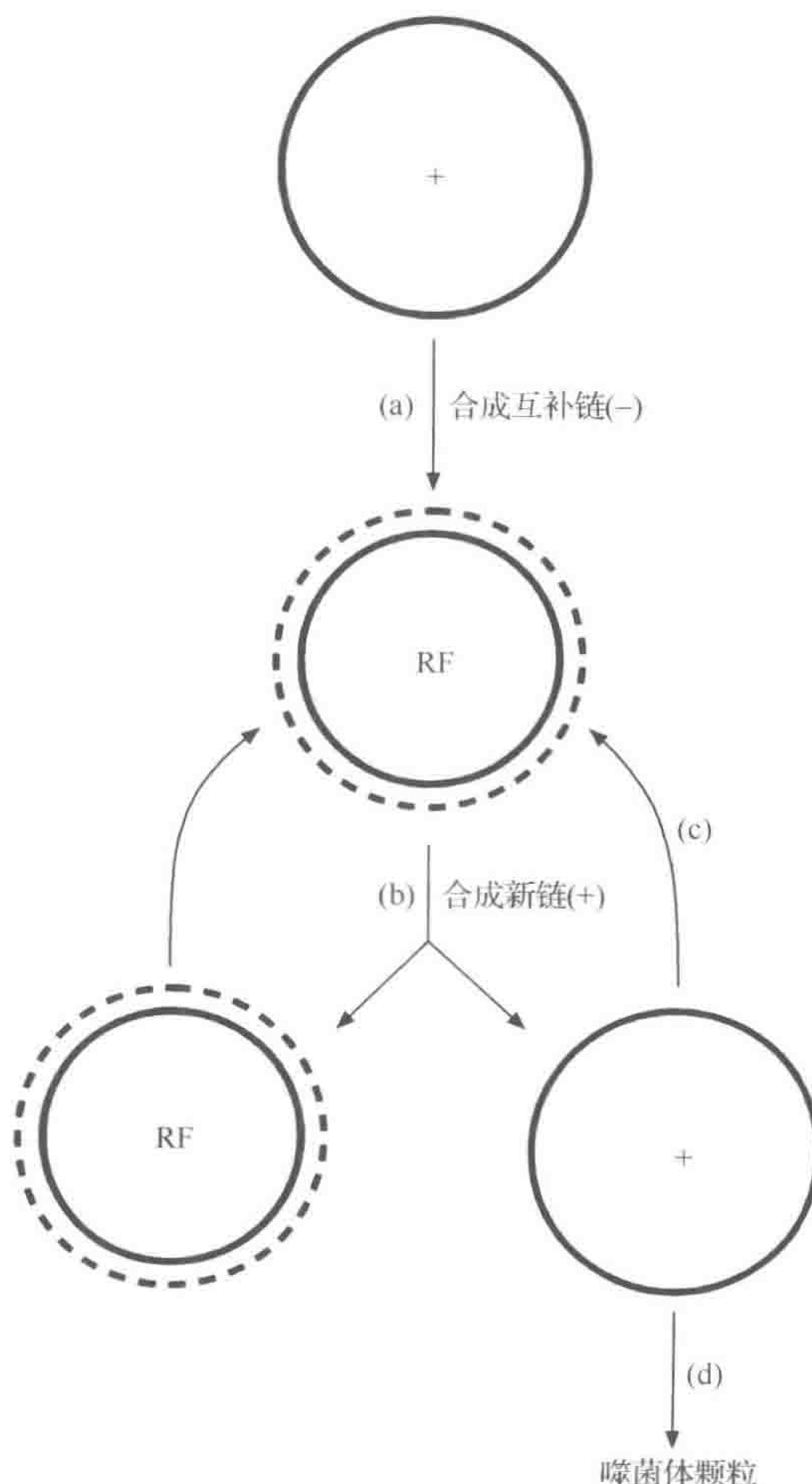


图 4.5 单链噬菌体的复制

DNA 进入细胞时是单链的，随后转变为双链的复制型（replicative form，RF）(a)；其中一条链被拷贝(b)，产生更多的单链分子，这些单链分子又转变成双链的复制状态(c)；最后(d)单链 DNA 被包装成噬菌体颗粒

链转变成双链 DNA 分子(复制型,RF)[步骤(a)];这条"负"链被转录用于噬菌体蛋白质的合成。由于噬菌体 DNA 的两条链是分开复制的,该复制方式与染色体 DNA 的复制方式不同。这条"负"链为下一步"正"链的复制提供模板[步骤(b)];步骤(b)合成的"正"链再合成互补链"负"链,形成双链 DNA 复制产物[步骤(c)]。虽然都以合成双链 DNA 结束,但是两条链分开复制、且过程不同。一些类型的质粒的复制方式与它们也非常相似,这将在第 5 章进行进一步讨论。

在 DNA 复制的同时,噬菌体蛋白质合成并组装到空的噬菌体颗粒前体中,其结果就是导致 DNA 复制途径的改变。新生成的"正"链不转变成双链的 RF,而是被空的噬菌体衣壳前体捕获,形成噬菌体的成熟颗粒[步骤(d)]。

4.2.2 M13

丝状噬菌体,如 M13,代表了单链噬菌体的另一种类型。因为它们通过依附于 F 质粒特异的菌毛末梢感染细胞(见第 5 章),又被称为"雄性专一"噬菌体。与其他毒力细菌病毒不同的是,它们不裂解宿主细胞,而是从感染的细胞中分泌出噬菌体颗粒,宿主细胞仍能继续生长和分裂。虽然它们同样会给细菌菌落造成噬菌斑样的损伤,但是只是抑制宿主细胞的生长而不是形成真正的噬菌斑。在感染宿主细胞后,通过合成其互补链从而形成双链 DNA 分子,这与 φX174 基本类似(图 4.5)。接下来,如图 4.5 所描述的那样,合成单链 DNA,转换成双链进行复制。

这类噬菌体的另一个特征是 DNA 不被包装到空的噬菌体头部前体中。随着复制的循环进行,合成的某一种蛋白质能结合到单链 DNA 上,使其生成噬菌体颗粒,通过靶向到宿主细胞膜,从细胞分泌出去,同时,在其通过细胞膜的过程中,噬菌体衣壳蛋白围绕其聚合。复制型 DNA(RF DNA)仍然停留在宿主细胞内,合成更多的噬菌体 DNA,进而产生更多的噬菌体颗粒。

4.3 RNA 噬菌体:MS2

MS2 是另一种"雄性特异性"噬菌体。它是一种二十面体的 RNA 噬菌体,与 M13 不同,MS2 依附于 F 质粒特异的菌毛侧部进入宿主细胞内。MS2 结构非常简单,只有 3600 个核苷酸,仅仅编码 3 种蛋白质:衣壳蛋白、成熟蛋白和复制酶。噬菌体颗粒中的 RNA 是一条编码链,既是复制模板,也是 mRNA。调控发生在翻译水平,主要由大量的 RNA 二级结构介导。

MS2 的复制需要 RNA 介导的 RNA 聚合酶,细菌细胞中通常不存在这种酶。因此,MS2 只有在 MS2 复制酶基因被翻译后才能开始复制。复制酶首先合成病毒链(正链)的互补链(负链),然后利用"负链"合成大量病毒 RNA。衣壳蛋白围绕"负链"RNA 聚集,当噬菌体颗粒在宿主细胞中达到一定数目后(5000~10 000/细胞),可能是由于引起了宿主细胞的机械性损伤,噬菌体颗粒得到释放。

4.4 双链 DNA 噬菌体

较大的噬菌体,如 T4 和 λ,均为双链 DNA 噬菌体。由于噬菌体颗粒内的 DNA 是线性的,结构较为复杂。细菌细胞内的 DNA 通常是环状的并非偶然。在常见的机制下,线性 DNA 分子的末端不能被复制,环状 DNA 分子的复制就不会遇到这样的问题。(一些细菌具有线性质粒,甚至线性染色体,并且已经进化到其他的方式以确保末端能被复制;见第 5 章)。因此,线性的噬菌体 DNA 在被复制前,通常转变成环型,这将在介绍 λ 噬菌体时详细描述。T4 噬菌体的复制机制不同于 λ 噬菌体(如下节)。

该类噬菌体复制的溶源模式的一般特征与概述所描写的相似。DNA 复制的最终结果是得到一条极长的 DNA 分子,比单条噬菌体基因组长数倍。与此同时,噬菌体尾巴和空的噬菌体头部被组装;多元化的长噬菌体 DNA 分子被剪切成正确的长度后被包装入空噬菌体头部,然后尾巴黏附并使靶细胞裂解。尽管 λ 和 T4 噬菌体的这个过程大致相似,但其差异也很明显,值得进行更深入的讨论。

4.4.1 T4 噬菌体

复制和包装

T4 噬菌体是一种感染大肠杆菌的毒力噬菌体,含有一条长约 165kb 的线状 DNA 分子,比含有全部噬菌体基因组所需的碱基序列稍长,因为表现出末端冗余或末端重复:位于一个 DNA 分子序列一端的一段长约 1600bp 的序列,并在该分子的另一端按同样的方向重复。正是由于这些重复序列的存在,通过末端重复序列之间的基因重组,促进了长的、由多条 DNA 组成的噬菌体 DNA 串联体复制产物的形成(图 4.6),使得整个 DNA 分子以线性结构被完整复制,不需环化,也不丢失末端序列。

其结果是产生了大量的比需要填充到单个噬菌体头部长许多倍的线性 DNA 分子。这时,形成噬菌体颗粒所需的各种蛋白质已经合成,并被组装成噬菌体尾巴和中空的头部。然后,在头部里面,DNA 被紧紧地卷成螺旋形,直到头部被填满为止,这样,DNA 被包装进入头部。DNA 链被剪切并且剩余的未结合的 DNA 被包装到另一个噬菌体头部。这种系统被称为“头满装”机制。

由于需要被包装进入头部的 DNA 数量比 T4 基因组完整长度多,这种“头满装”机制导致了末端冗余。如果包装在 A 点起始,继续穿过基因组远末端,最后加上 A 区域第二段。另外,这种机制造成的另一种结果就是,在不同的噬菌体颗粒中,基因次序呈现环状排列现象;也就是说,虽然碱基按环状图谱编排,排列顺序通常是一样的,一个噬菌体线性 DNA 的起点和终点与其在另一个噬菌体颗粒的起点和终点是不同的。在上面的例子中,如果第二个噬菌体颗粒的起点和终点在 B 区域,则第三个噬菌体的起点和终点在 C 区域,以此类推(图 4.6)。

噬菌体生长控制

在前面概述中提到,大多数噬菌体的一个典型的特征是:在感染细胞后,特异的一组

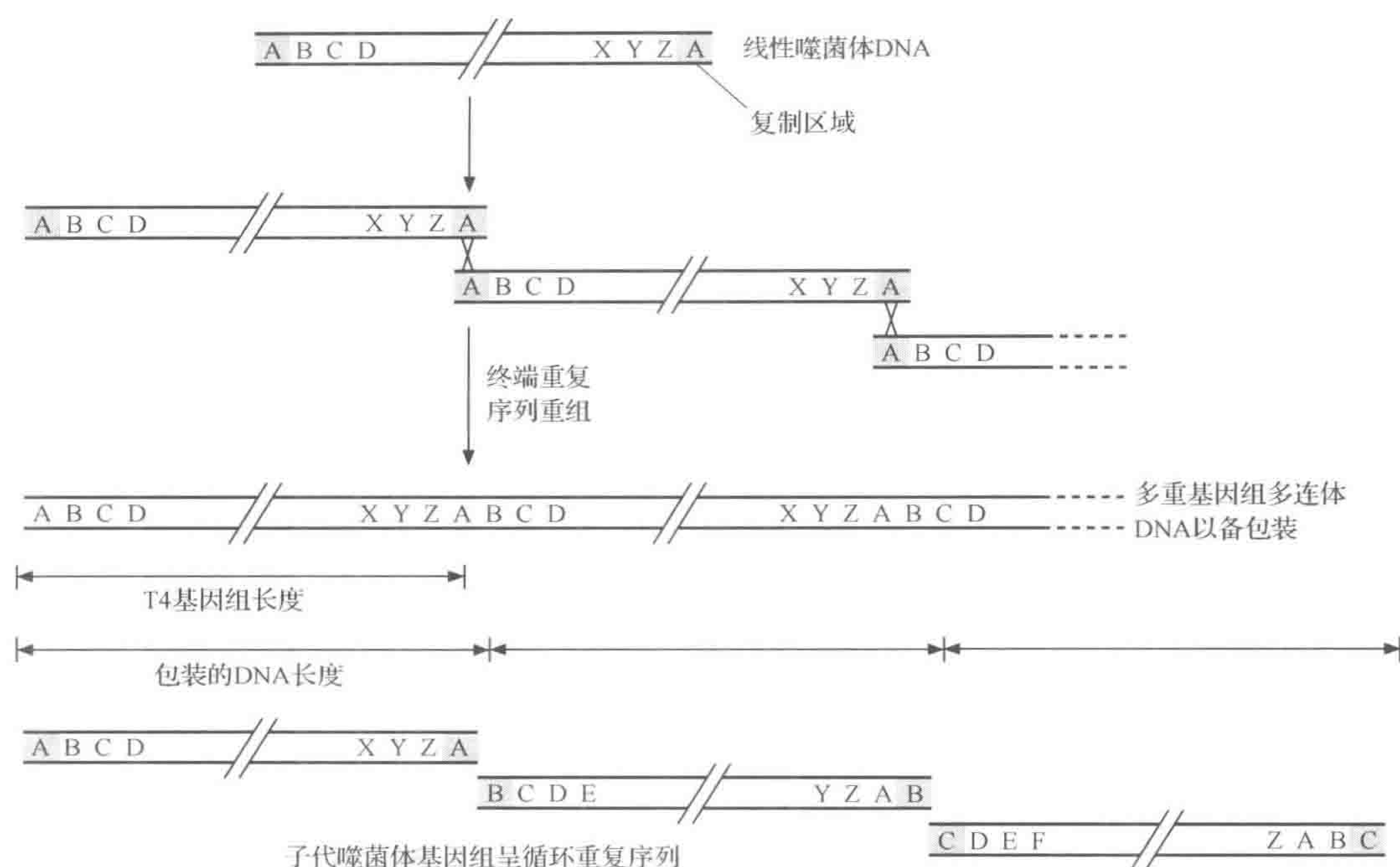


图 4.6 T4 噬菌体的复制

噬菌体线状 DNA 的复制伴随着末端重复的重组过程，产生多重基因组多连体 DNA 包装成噬菌体颗粒

基因在不同时期表达。最重要的分界线为早期基因和晚期基因，早期基因是指噬菌体感染宿主细胞后即利用宿主的酶合成蛋白质的基因；晚期基因的表达需要利用一种噬菌体自身基因编码合成的蛋白质，这种蛋白质是一种早期基因编码的产物。

关于 T4 噬菌体，早期基因的启动子能被宿主 RNA 聚合酶识别，从而得到表达。第二组基因，被称为(quasi-late)准晚基因(或中期基因)，在感染后稍晚被表达；这些基因同样由宿主 RNA 聚合酶转录，但是启动子不同。反之，两个噬菌体基因编码的蛋白质被用来帮助 RNA 聚合酶结合到 DNA 分子上。总体来说，这两组基因包含了噬菌体 DNA 的复制和调节后期基因合成所需的信息，这些后期基因自身携带有产生噬菌体颗粒的信息。当一个新的可以改变宿主 RNA 聚合酶的特异性的 σ 因子产生后，晚期基因被激活(见第3章)，但是另外的调控机制，包括衰减子的存在，同样被用来调控 T4 基因的暂时表达。

另外一种在基因表达中具有相似的、特异开关的噬菌体是枯草芽孢杆菌噬菌体 SPO1，其早期基因被宿主 RNA 聚合酶转录(使用前体 σ 因子)。其中一个早期基因(基因28)编码 σ 因子(gp28)，该因子是一组噬菌体基因，即中期基因的分子开关。这些中期基因编码的其中两种产物(gp33 和 gp34)可反过来结合于 RNA 聚合酶核心部位，从而使晚期基因激活。

4.4.2 λ 噬菌体

复制和包装

与 T4 噬菌体相同，λ 噬菌体颗粒也含有线性双链 DNA。然而，与 T4 噬菌体不同的

是，λDNA 既没有末端冗余，也没有环状排列。所有噬菌体颗粒中 DNA 末端是高度一致的。这反映了 λ 噬菌体和 T4 噬菌体之间不同的复制和包装过程。

当 λ 噬菌体感染 *E. coli* 时，线性 DNA 被注入靶细胞中，并借助于线性 DNA 5′端的 12 个不配对碱基组成的黏性末端(黏端)，转化成环状(见图 4.8)。两端的序列(*cos* 序列)之间互补，并呈黏性，也就是说，一端的序列趋向于和另一端的序列相互配对，然后形成(非共价键结合)环状的 DNA 分子。靶细胞内的 DNA 连接酶将环状 DNA 的缺口快速地连接起来，从而形成共价闭合环状 DNA 分子。

λ 噬菌体的 DNA 最初以 θ 形式复制（因为复制时的空间结构与希腊字母 θ 很相似而得名）。这与质粒等其他环状 DNA 分子的复制方式基本相似。到了感染的晚期，控制滚环式复制机制的开关被启动，从 θ 形式转变为滚环式复制，形成了长多连体线性 DNA 分子(图 4.7)。和 T4 一样，到了这个时期，噬菌体尾巴和中空的头部已经合成，然而，其包装机制不同。

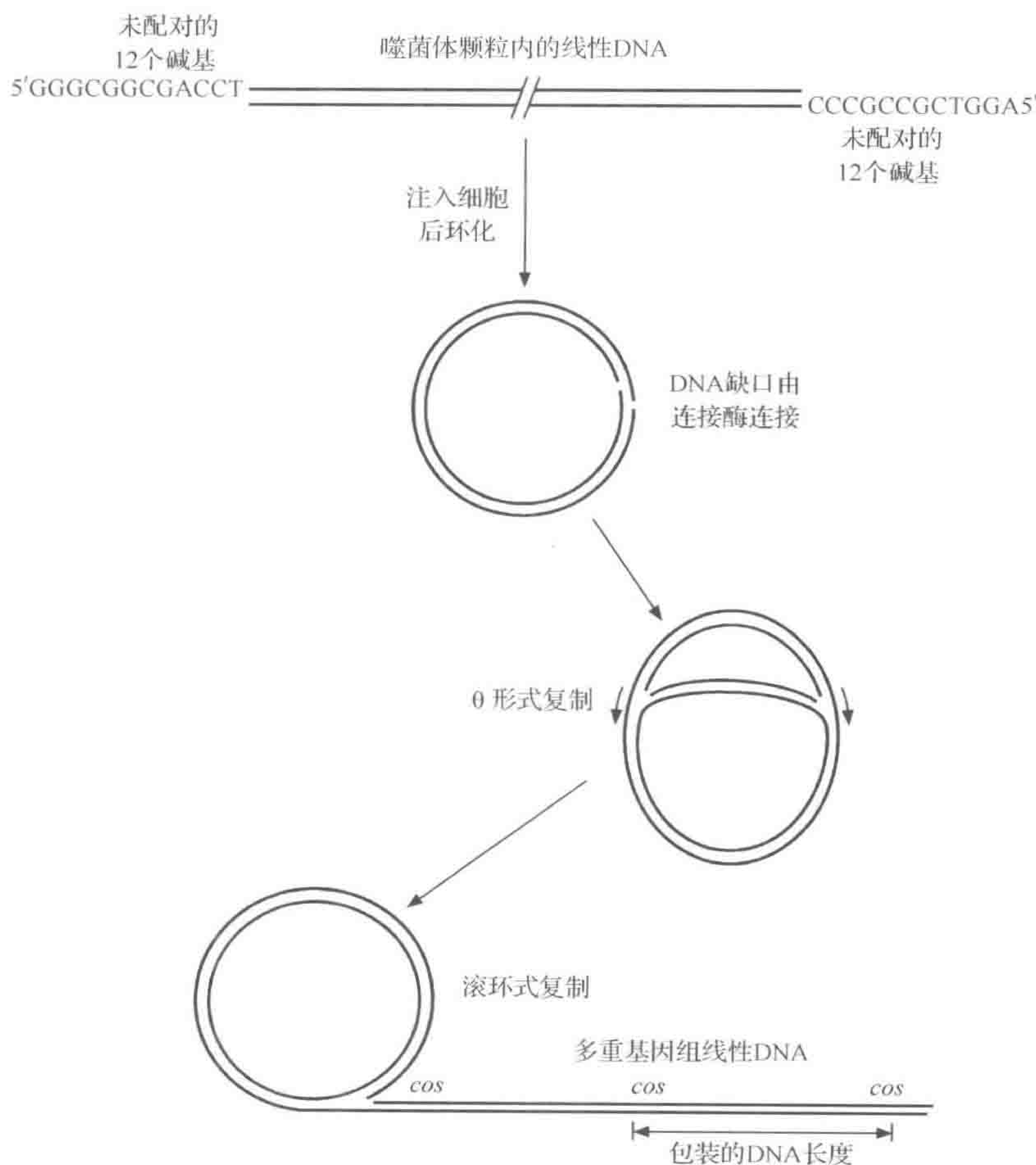


图 4.7　λ 噬菌体的复制

感染后，DNA 黏端连接形成环状分子，环状分子复制(θ 形式)产生更多的环状 DNA。随后，转变成滚环式复制，产生多重基因组线性 DNA 以包装成噬菌体颗粒

关于λ噬菌体，一段特殊序列的位置，*cos*位点决定了所包装DNA的大小。某种蛋白质吸附于噬菌体头部，在长多连体线性DNA分子上识别一个*cos*位点，从而启动DNA的包装机制。当到达下一个*cos*位点时，这个蛋白质在每个*cos*位点处剪切DNA。这种剪切是不对称的，也就是说，两条单链被剪切的碱基位置不是对应的。两条链的碱基缺口有12个碱基的距离，在每一个包装DNA的两端产生了一段12个不配对碱基的黏端序列(图4.8)。然后封闭头部，加上前体尾巴，宿主被裂解，产生的新的噬菌体颗粒得到释放。

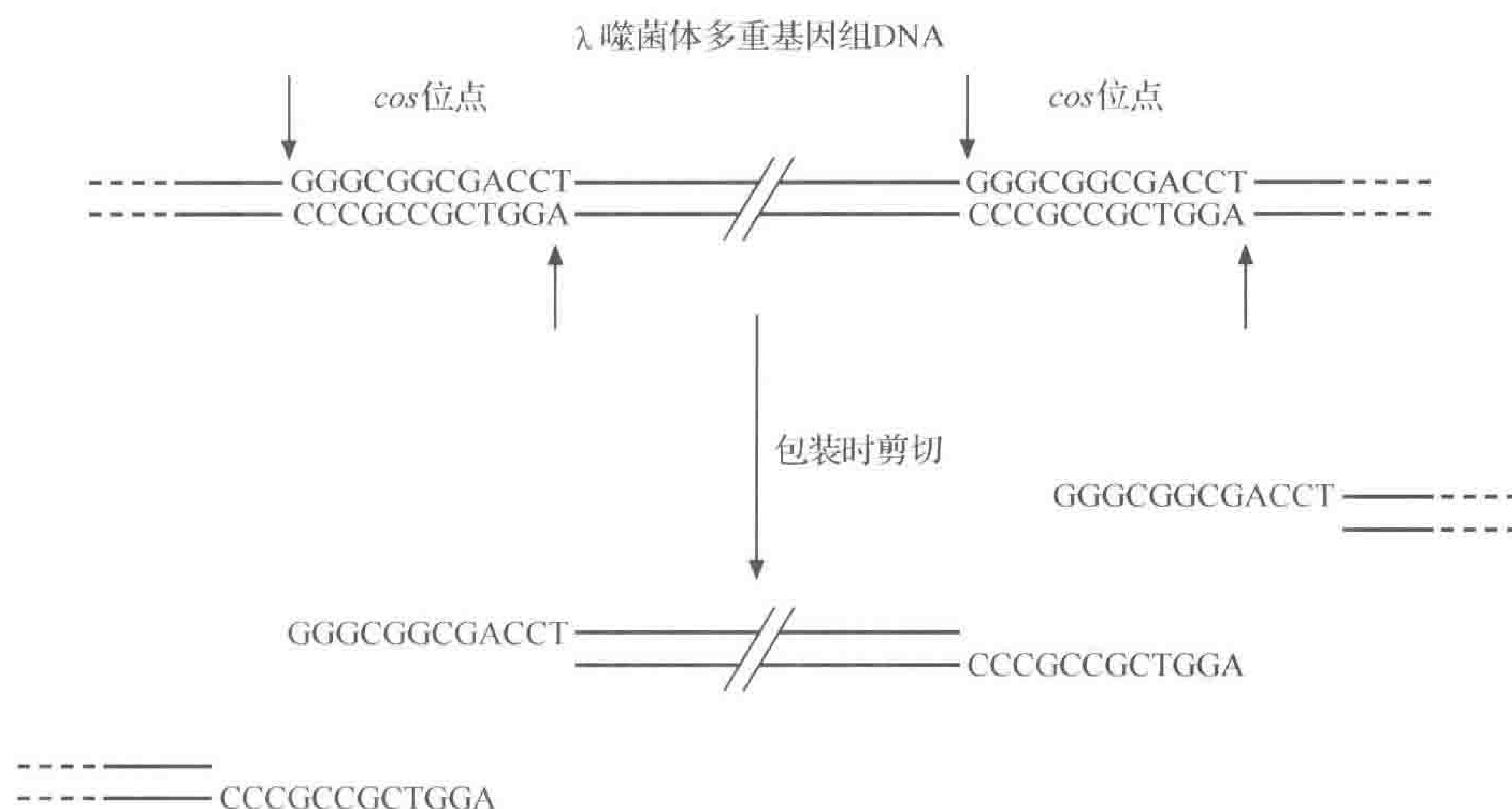


图4.8　λ DNA在*cos*位点被剪切

这种包装机制在λ噬菌体作为基因克隆载体的实际应用中具有重要意义。因为两个*cos*位点之间的距离决定了包装的DNA数量，DNA缺失可能意味着不足的DNA被包装进入噬菌体头部来维持它的物理完整性。另外，向λ基因组DNA中加入DNA片段可能会使两个*cos*位点之间的距离超出噬菌体头部的容量，这样的DNA不能被包装。因此，λ噬菌体具有包装限制；两个*cos*位点之间的距离必须为野生型序列的75％～105％(即为37～52kb，λ噬菌体野生型基因组全长为48.5kb)。这部分内容将在第8章进行进一步讨论。

溶源

λ噬菌体是一种温和型噬菌体，在感染细菌细胞后，可与宿主细胞形成溶源关系(图4.9)。当λ噬菌体或其他温和噬菌体感染细菌培养物时，在部分细胞中噬菌体会进入裂解周期，同时，其他宿主细胞会变成溶源菌。在溶源状态时，噬菌体DNA维持在前噬菌体状态，特异地抑制蛋白质的作用，阻止了裂解循环。

追踪溶源通路发现，代替θ形式复制的是，λ噬菌体DNA通过噬菌体上的特殊基因(*attP*)和细菌染色体基因(*attB*)重组(图4.10)整合到宿主的基因组中。这两段基因序列(*attP*和*attB*)均有一段短的完全相同的区域，即核心序列(15bp)。基因重组需要λ噬菌

体基因 *int* 的编码产物(整合酶)。和普遍的基因重组机制明显不同(见第 2 章),不需要一段延伸的同源性区域。因为这类基因重组只发生在特异的靶位点,是位点特异性重组中的一个例子。

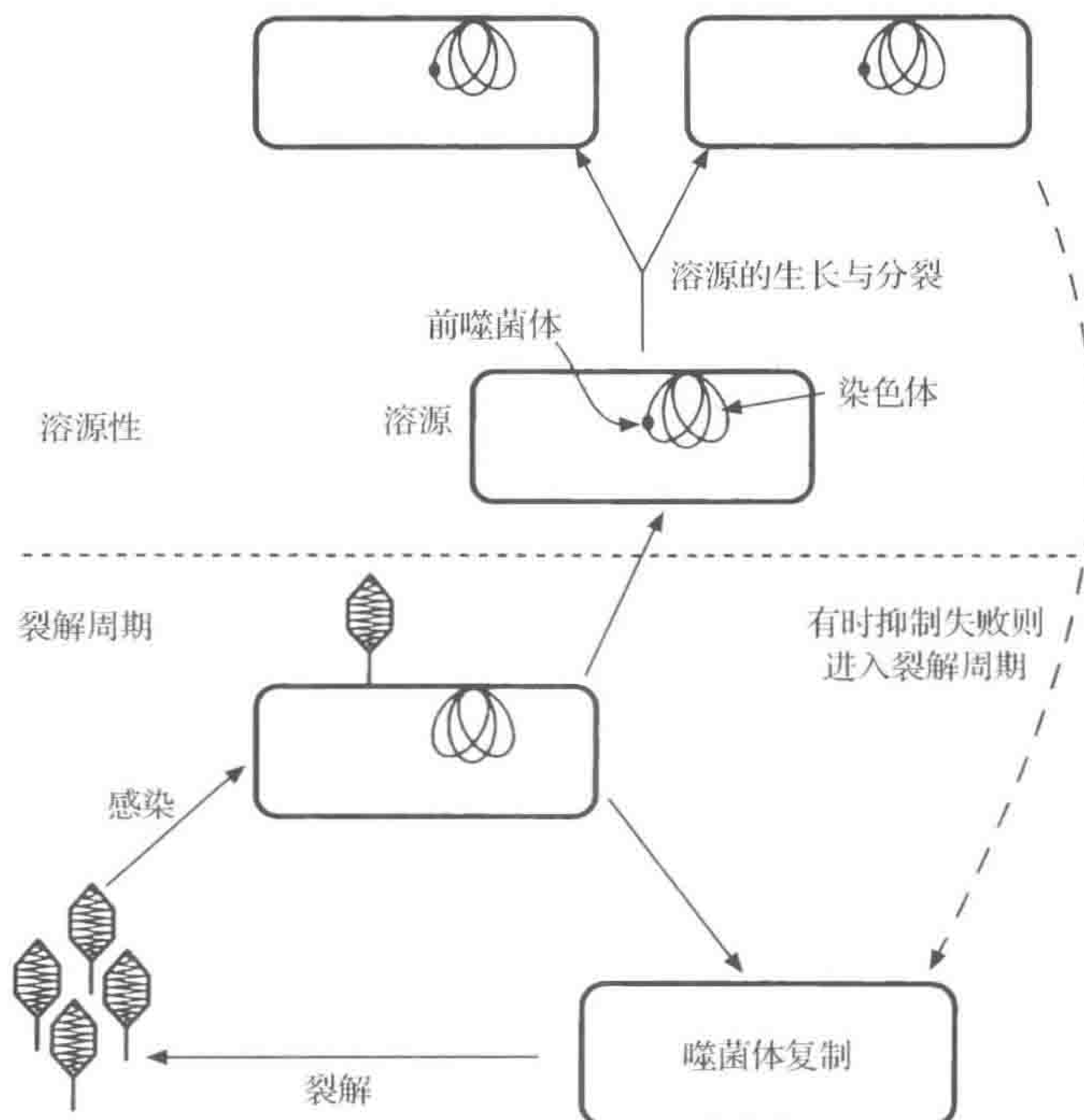


图 4.9 裂解周期和溶源性:温和噬菌体的选择

感染后,温和噬菌体可产生溶源性,是不同于裂解周期的另一个选择。本图中,前噬菌体与细菌的染色体进行了整合,但对某些噬菌体,前噬菌体以质粒存在

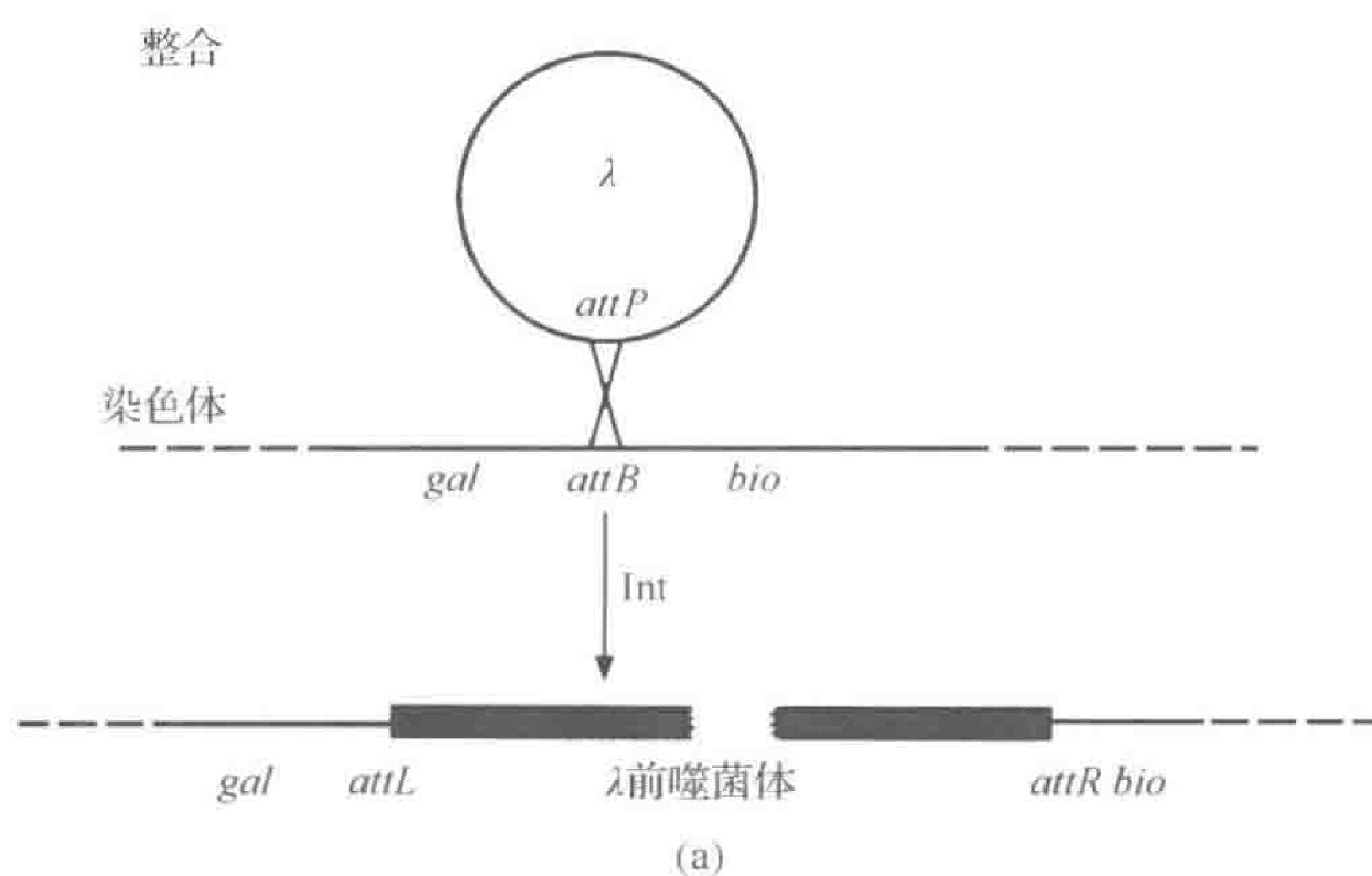
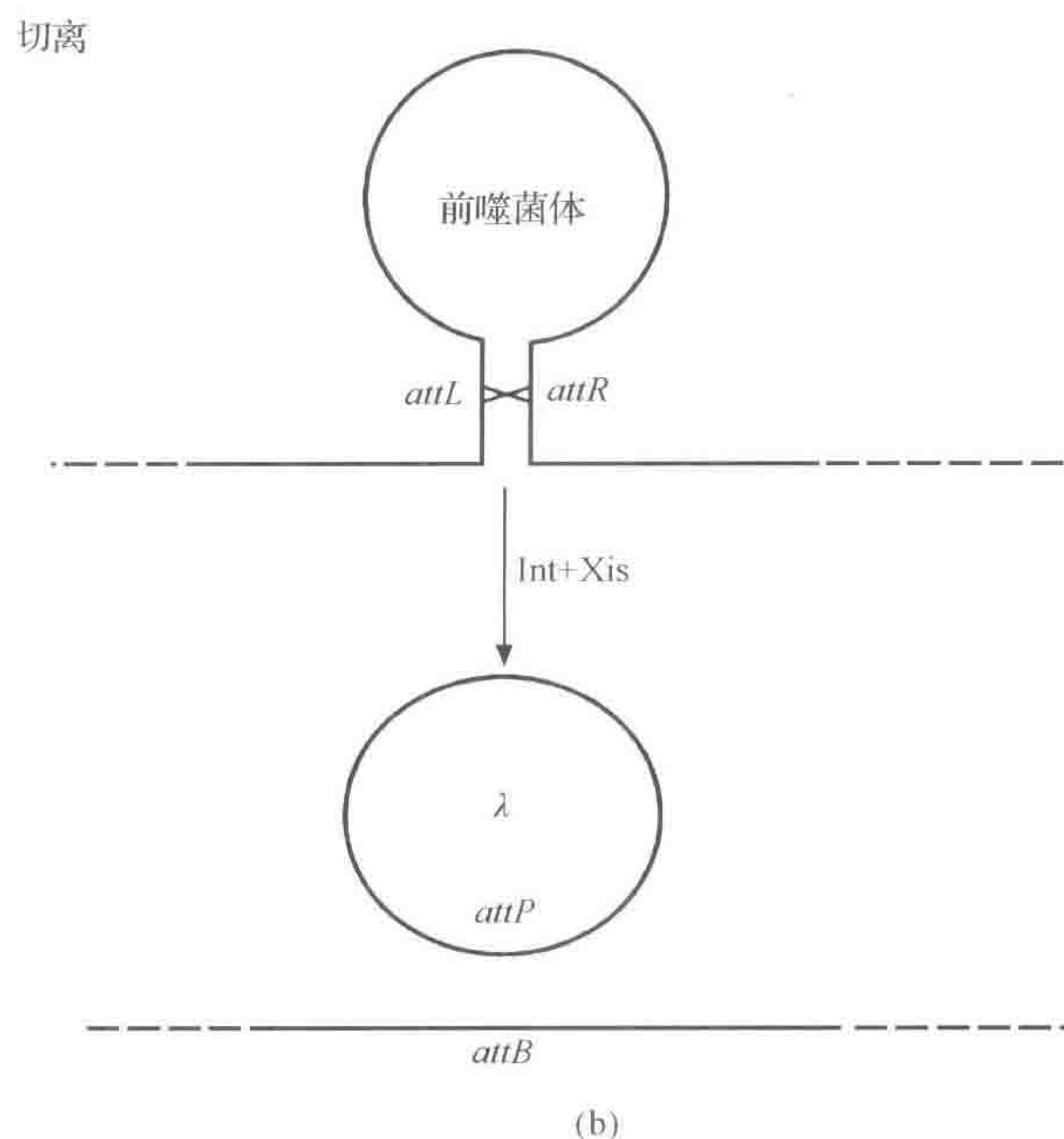


图 4.10 λDNA 的整合和切离



(b)

图 4.10 λDNA 的整合和切离(续图)

(a) 在 Int 蛋白的参与下,在特异的重组位点发生整合;(b) 切离时同时需要 Int 和 Xis

因为 *attP* 和 *attB* 序列是不同的(除中心的核心区域外),基因重组后产生两条与 *attP* 和 *attB* 均不相同的序列,分别位于核心区域的两端。因位于前噬菌体基因组的左右两侧,称为 *attL* 和 *attR*(图 4.11)。完成位点特异性重组的整合酶,能特异性识别 *attP* 和 *attB* 位点,而不会催化 *attL* 和 *attR* 之间的基因重组。因此,整合酶不能启动会导致前噬菌体切出的逆向重组,从而确保了在维持溶源状态中整合作用是不可逆的。溶源状态的破坏会导致前噬菌体进入裂解周期(由溶源细菌诱导),另一种基因(*xis*)的表达产物与 Int 的相互作用改变了 Int 的特性,使之被启动,导致一个完整无缺的染色体和一个环状

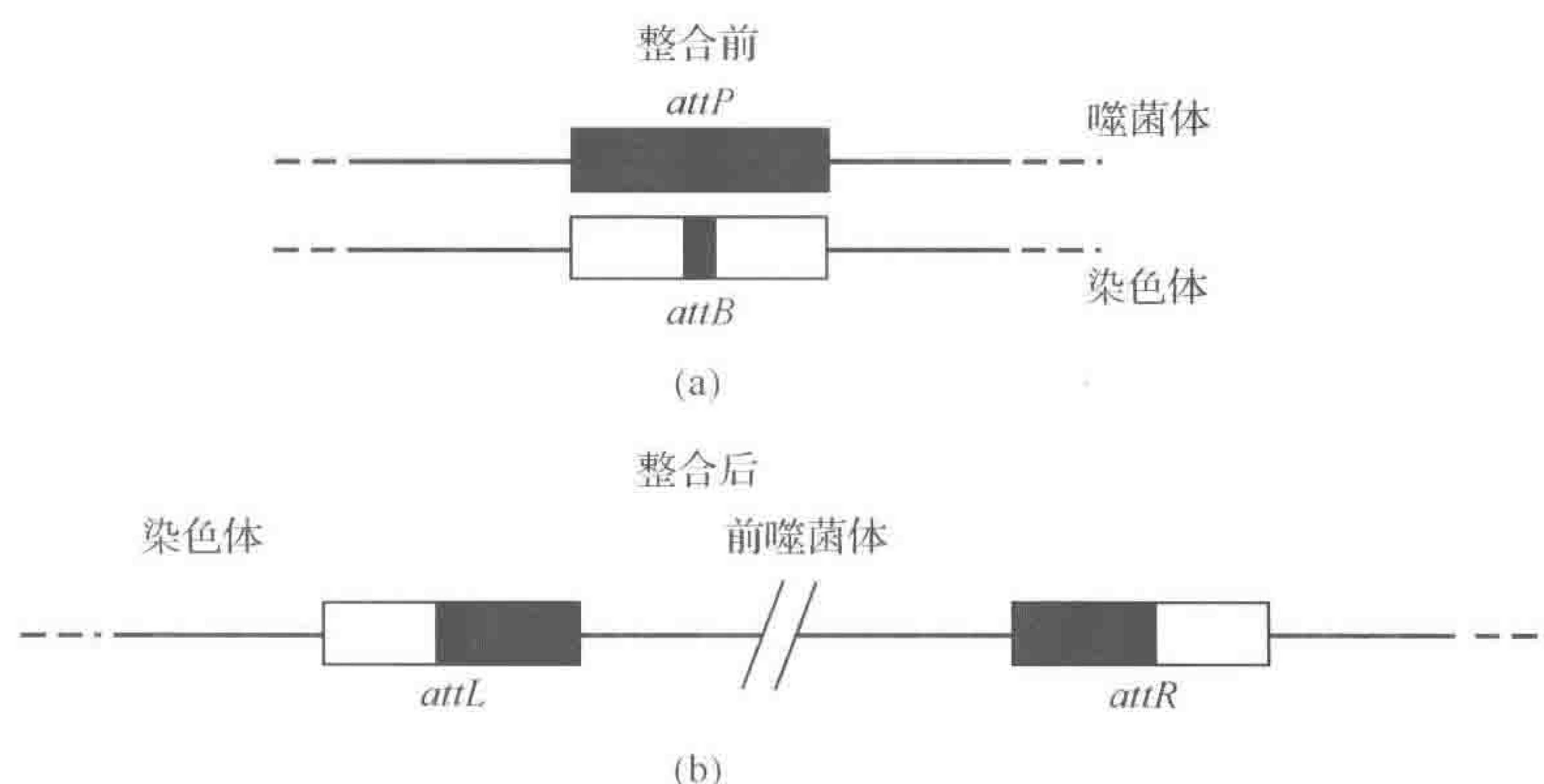


图 4.11 λ 噬菌体的连接位点结构

(a) 整合需要位于 *attP*(在噬菌体上)和 *attB*(在染色体组上)之间的特异性重组位点;(b) 整合完成后,在前噬菌体(*attL* 和 *attR*)两个末端具有与 *attP* 和 *attB* 相同的核心序列(黑色),但其余两侧序列是不同的

噬菌体 DNA 分子的形成(图 4.10)。

需要强调的是,整合不是溶源性的基本特征。一些不能整合的 λ 噬菌体的突变体也可以使宿主变成溶源菌,前噬菌体以环状噬菌体 DNA 游离在宿主染色体外的方式存在,在功能方面相当于质粒。这些 λ 溶源菌通常是不稳定的,但是也有一些温和噬菌体,如 P1,形成了稳定的溶源菌,前噬菌体就是质粒。在后面的章节将 P1 作为一个克隆载体进行讨论。

4.4.3 λ 噬菌体的裂解和溶源调控

因为它为充分了解噬菌体的不同调控机制提供了例子,对 λ 噬菌体做更详细深入的研究是值得的。尤其是可以寻找暂时地控制裂解周期(即不同位点基因的连续表达是怎样完成的)和溶源性的方法,后者包括了决定个体噬菌体是否进入裂解或溶源状态(裂解/溶源周期)开关的性质。

裂解周期的瞬时控制

图 4.12 概括了 λ 基因组的主要特征,从图中可以看出其基因组按功能组排列。早期

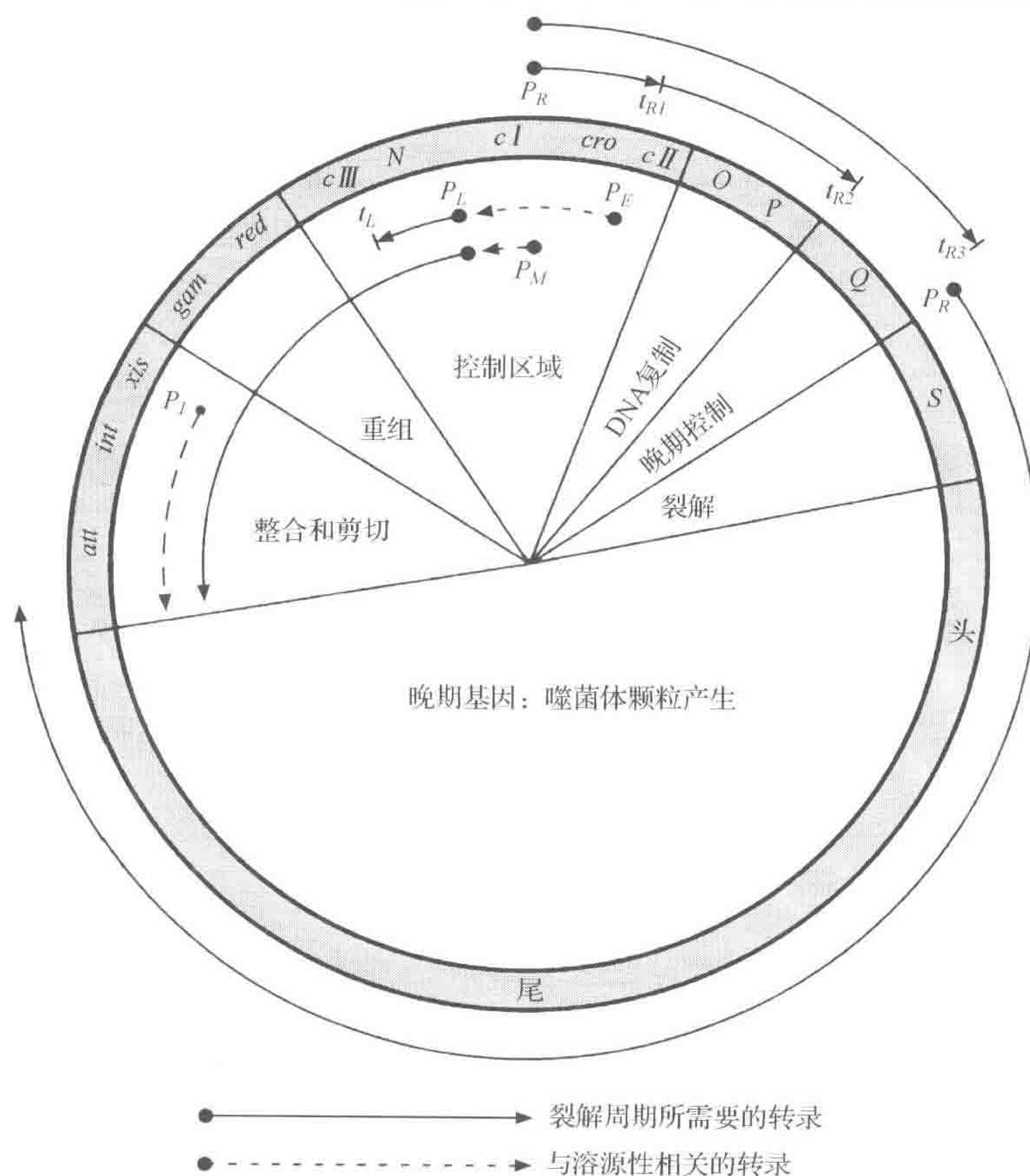


图 4.12 λ 噬菌体:遗传图和转录组成

基因，主要和 DNA 复制有关，距离控制区域较近，早期基因的转录本大多起始于这里。在感染后期，发现晚期基因的表达产物，主要合成构成噬菌体颗粒头部和尾巴的蛋白质。图 4.13显示了放大后的控制区域。

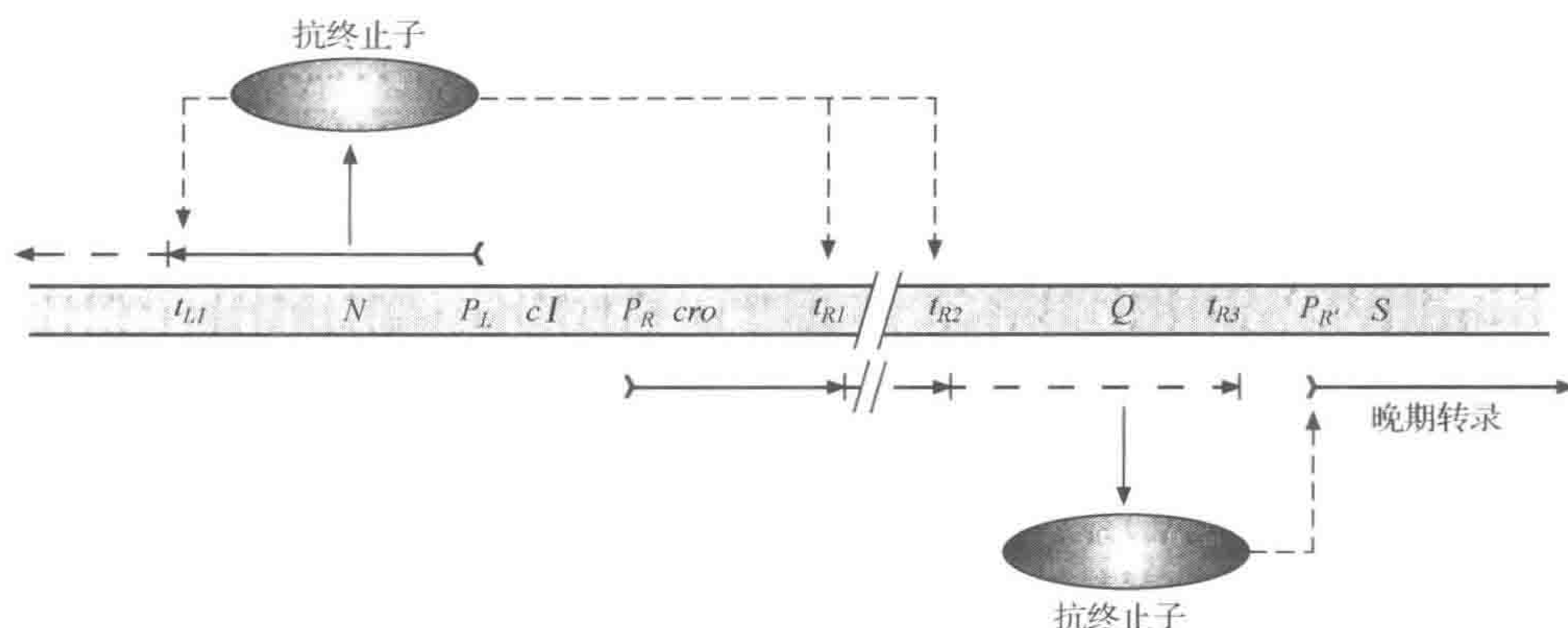


图 4.13 λ 噬菌体裂解转录的调控

早期转录起始于 P_L和 P_R，终止于 t_{L1}和 t_{R1}/t_{R2}。N 基因产物具有抗终止活性，并使早期的延迟基因表达，向右侧转录终止于 t_{R3}。Q 基因产物也是抗终止子，并使晚期基因从 $P_{R'}$处表达

λ 噬菌体感染细菌后不久，翻译主要起始于两个能被 *E. coli* RNA 聚合酶有效识别的主要启动子，分别是左向启动子 P_L和右向启动子 P_R。右向转录起始于 P_R，终止于转录终止子 t_{R1}和 t_{R2}(t_{R1}是弱终止子，一些转录本可以跳过，继续向后翻译)；左向翻译终止于 t_{L1}。因此，只有极少数基因可立即表达，其中包括基因 *N* 和 *cro*。

基因 *N* 的表达产物具有特异的抗终止活性，即它可以使翻译通过终止子 t_{R1}、t_{R2}和 t_{L1}。N 蛋白结合于终止子上游 DNA 的特殊位点，和 RNA 聚合酶发生相互作用，从而通过这些转录终止子位点。一组稍远的基因(“延迟早期”基因)就包括在这些延长的转录产物中。右向转录到足够远，即一些不受 N 蛋白活性影响的 t_{R3}强终止子处，右向转录的延迟早期基因涉及 DNA 的复制；尽管左向转录的基因对裂解繁殖来说不是很重要，但是有利于噬菌体的繁殖，如 *red* 和 *gam* 等。(为了使描述合理、简单，将忽略左向转录的发生)。

延迟早期基因(*Q*)的产物也是一个抗终止子，使来自于基因 *Q* 和基因 *S* 之间的启动子 $P_{R'}$的晚期基因转录。这导致了一个包括噬菌体的头和尾的全部蛋白质，以及为其包装和成熟所必需的其他产物信息的很大的 mRNA 分子的产生(见图 4.12)。

在以上事件发生的同时，早期基因 (*cro*) 之一的产物水平升高，该产物成为早期启动子 P_L和 P_R的阻遏物，并且阻止了不再需要的早期和延迟早期转录物的进一步合成。Cro 对于 P_L和 P_R的阻遏也在裂解/溶源开关中发挥了关键作用。

溶源性的调控

以上对裂解周期调控的描述忽略了 λ 复制的另一种模式，即通过溶源态的形成进行复制。当细菌处于溶源态时，*c* Ⅰ 是关键基因(*c* 表示清晰，与温和噬菌体产生混浊的噬菌斑相反，因为突变体不能成为溶源菌，因此会形成清晰的噬菌斑)。该基因编码的阻遏蛋

白 c Ⅰ能阻断噬菌体的裂解通路,从而使噬菌体与宿主细胞形成稳定的关系。将在下面讨论该阻遏蛋白的作用方式。

启动子 P_E调控 c Ⅰ基因表达(发生阻遏),而 P_E位于 *cro* 基因右侧(图 4.14)。但是,这个启动子只有在 c Ⅱ表达产物存在的情况下才发挥功能,因此 c Ⅱ基因是 P_E启动子的正调控子。另一个基因 c Ⅱ的产物对 c Ⅱ基因表达产物起到稳定作用(c Ⅱ和 c Ⅲ的表达产物对蛋白水解酶比较敏感,易降解,因此它们的寿命很短)。这些基因分别从 P_R和 P_L启动子的右侧和左侧开始表达,但它们都越过了最初的终止位点,需要抗终止基因 *N* 的存在,因此在感染后不会立即表达。一旦这些基因表达的产物都存在,P_E启动子即被激活。

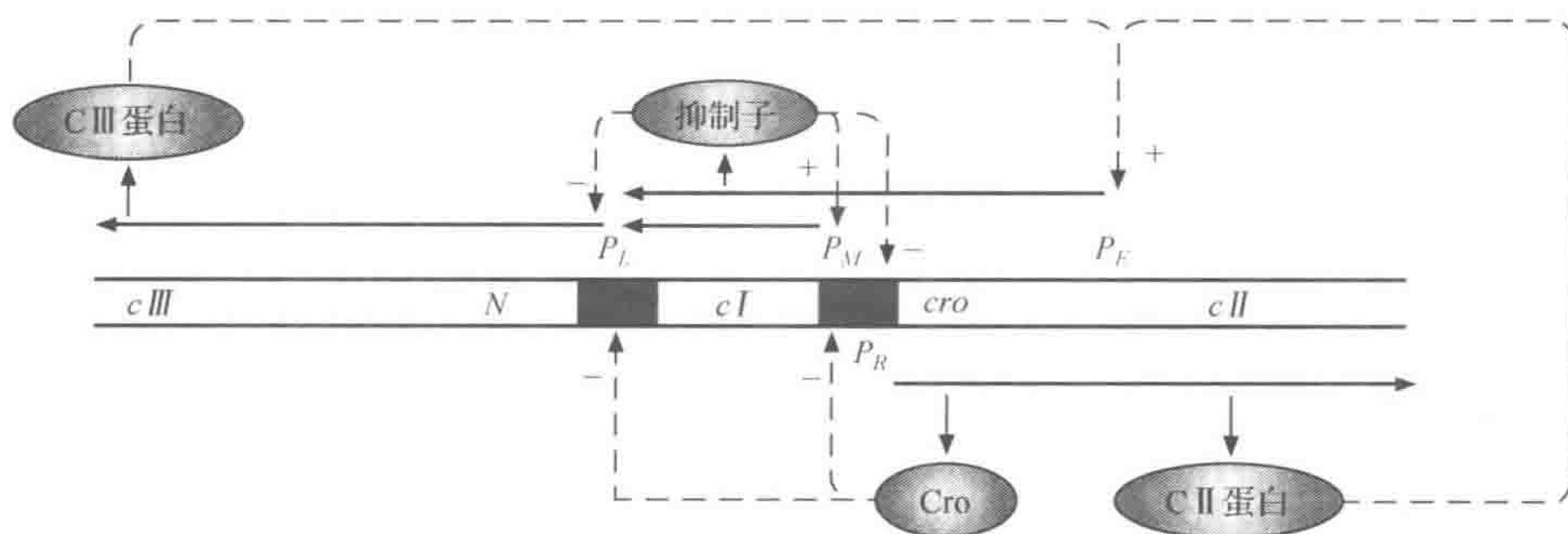


图 4.14 λ噬菌体裂解/溶源性的调控

CⅡ激活 P_E启动子,c Ⅰ抑制基因转录;CⅢ蛋白稳定 CⅡ蛋白,CⅠ蛋白是 P_L、P_R的主要阻遏物,并可激活 P_M合成更多的阻遏蛋白。Cro 蛋白是 P_L、P_R的阻遏蛋白,并可抑制 P_M

之后,阻遏蛋白(CⅠ)结合 λDNA 上的两个操纵子 O_L和 O_R,它们分别与 λ 早期主要的启动子 P_L和 P_R相邻。结合的一个结果是两个启动子被关闭,阻断了一些重要产物的进一步合成,特别是 *N* 基因和 *cro* 基因的产物。阻遏蛋白关闭启动子 P_L和 P_R的同时也关闭了 c Ⅱ和 c Ⅲ基因的表达,而后两者对于 P_E启动子的转录是必需的。由于要保持溶源状态,继续合成阻遏蛋白是必不可少的。这种情况通过另外一个与 O_R/P_R紧相邻的弱组成型启动子 P_M(维持启动子)在较低转录水平下实现。抑制子与操纵子的结合,不仅抑制 P_R的活性,而且激活了 P_M的活性,因此,CⅠ蛋白具有负向和正向的调节功能。

如前所述,Cro 蛋白也是通过结合 O_R位点而抑制 P_R。与结合阻遏蛋白相反,Cro 与 O_R的结合同时抑制 P_M启动子的活性。因此,Cro 蛋白通过两种方式来阻止 CⅠ阻遏蛋白的合成:通过关闭 c Ⅱ基因表达间接阻断 P_E启动子转录;与 O_R操纵子结合直接关闭组成型启动子 P_M。因此,可以把噬菌体裂解周期或溶源化(虽然很简单)看做 *cro* 和 c Ⅰ基因产物与 O_R位点竞争结合的结果。这些相互作用总结见图 4.15。下一节有关 CⅠ和 Cro 蛋白与 O_R操纵子结合的介绍,将有助于更好地理解这些关系。

为什么 λ 控制其生活周期的行为如此复杂?原因很简单也非常巧妙。噬菌体一旦进入某种生活状态,那么这种生活状态是自我强化并不可改变的。在感染的早期阶段,如果溶源性一旦形成,就快速而持续地产生阻遏蛋白,这使更多噬菌体的转录在无法回复进入

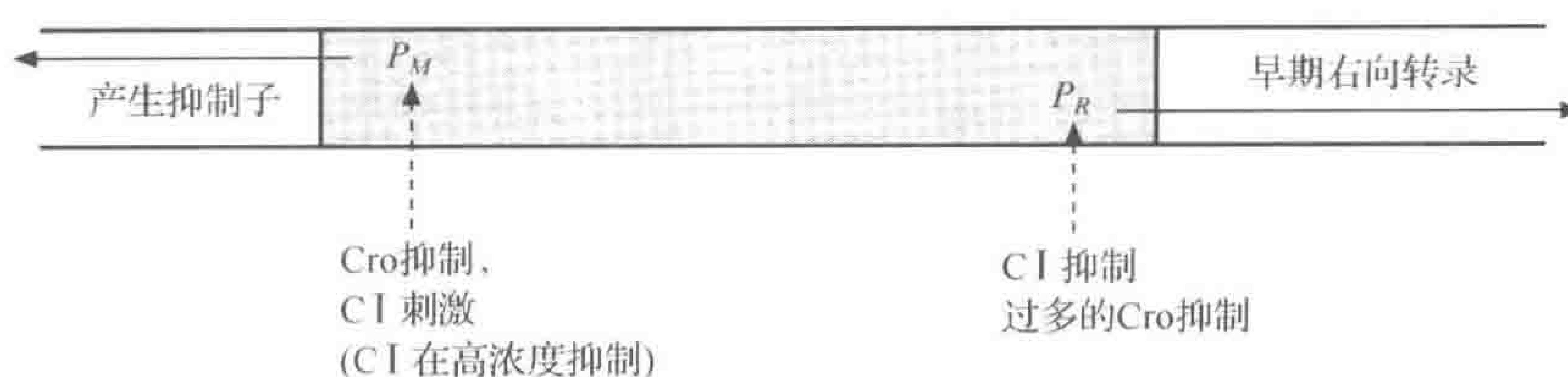


图 4.15 Cro 和 CⅠ对 P_M和 P_R的相互作用

裂解途径前被阻断。阻遏蛋白的产生通过高水平启动子 P_E来实现。另外,溶源性形成后,对阻遏蛋白持续合成的需求就保持在相对较低的水平,这可通过启动 P_M完成。因此,溶源性控制是自我强化的,CⅠ与 O_R结合阻止了其他基因的转录,但同时也确保了阻遏蛋白的持续合成。

如果阻遏蛋白从 O_R结合位点上移出(有时自发产生,也有可能是多种压力诱导原噬菌体形成的结果),不仅解除对 P_R的抑制,同时也阻断了 P_M。因此,阻遏蛋白的合成被阻断,无法替换已破坏的蛋白质。这也是自我强化的关闭机制。

阻遏蛋白和 Cro 蛋白与操纵位点结合的机制

在前面已经提到,如果操纵子 O_R仅被看做是一条简单的 DNA 序列,那么阻遏蛋白或 Cro 蛋白都能与它结合。由于这个系统提供了调控蛋白结合 DNA 的相关信息,因此,操纵位点和与其结合的这些蛋白质的特性就值得进行更加深入的研究。由于 O_L位点与它有很高的相似性,只考虑 O_R位点的结合机制。

事实上,O_R操纵子是由 3 个结构相似且毗邻的 DNA 区域组成:O_{R1}、O_{R2}和 O_{R3}(图 4.16)。CⅠ和 Cro 蛋白能与这 3 个位点中的任何一个结合,各自的亲和力不同,并且结合效果也不同。

阻遏蛋白与 O_{R1}结合的亲和力最强,并且这种结合具有一定的协同性,即阻遏蛋白结合 O_{R1}能激活第二个分子与 O_{R2}的结合。结合 O_{R2}的阻遏蛋白与 RNA 聚合酶作用,激活启动子 P_M开始转录,阻断 RNA 聚合酶与 P_R的结合。在正常的溶源状态下,O_{R3}不被结合,但当阻遏蛋白浓度处于较高水平下也与 O_{R3}结合,P_M的转录被阻断,直到阻遏蛋白的浓度降低到为饱和 O_{R3}所必需的水平为止。因此,除了阻遏蛋白外,不仅所有的基因表达被关闭,而且 *c*Ⅰ的表达水平也保持在一个较低的水平,这有利于维持噬菌体的稳定性,减轻宿主细胞的负担。

Cro 也能与 3 个位点结合,但亲和力相对阻遏蛋白有所不同:它首先与三者中的 O_{R3}结合,阻断 RNA 聚合酶结合组成型启动子 P_M,从而阻止噬菌体进入溶源状态。然后,当 Cro 水平升高,它与位点 1 和 2 结合;这时,噬菌体处于裂解状态,从 P_L转录产生大量的转录本。Cro 的作用是关闭不再需要的早期表达的基因,调控噬菌体颗粒合成所需的晚期基因表达。

由此,可以进一步完善 λ 噬菌体形成和调控的裂解生长/溶源模型。其关键在于CⅠ阻遏蛋白与 O_{R1}的结合阻断噬菌体裂解循环,相反,Cro 与 O_{R3}结合通过抑制启动子 P_M转录限制溶源状态的形成。但只要活化的 CⅡ蛋白存在,Cro 与操纵子的结合并不能阻断

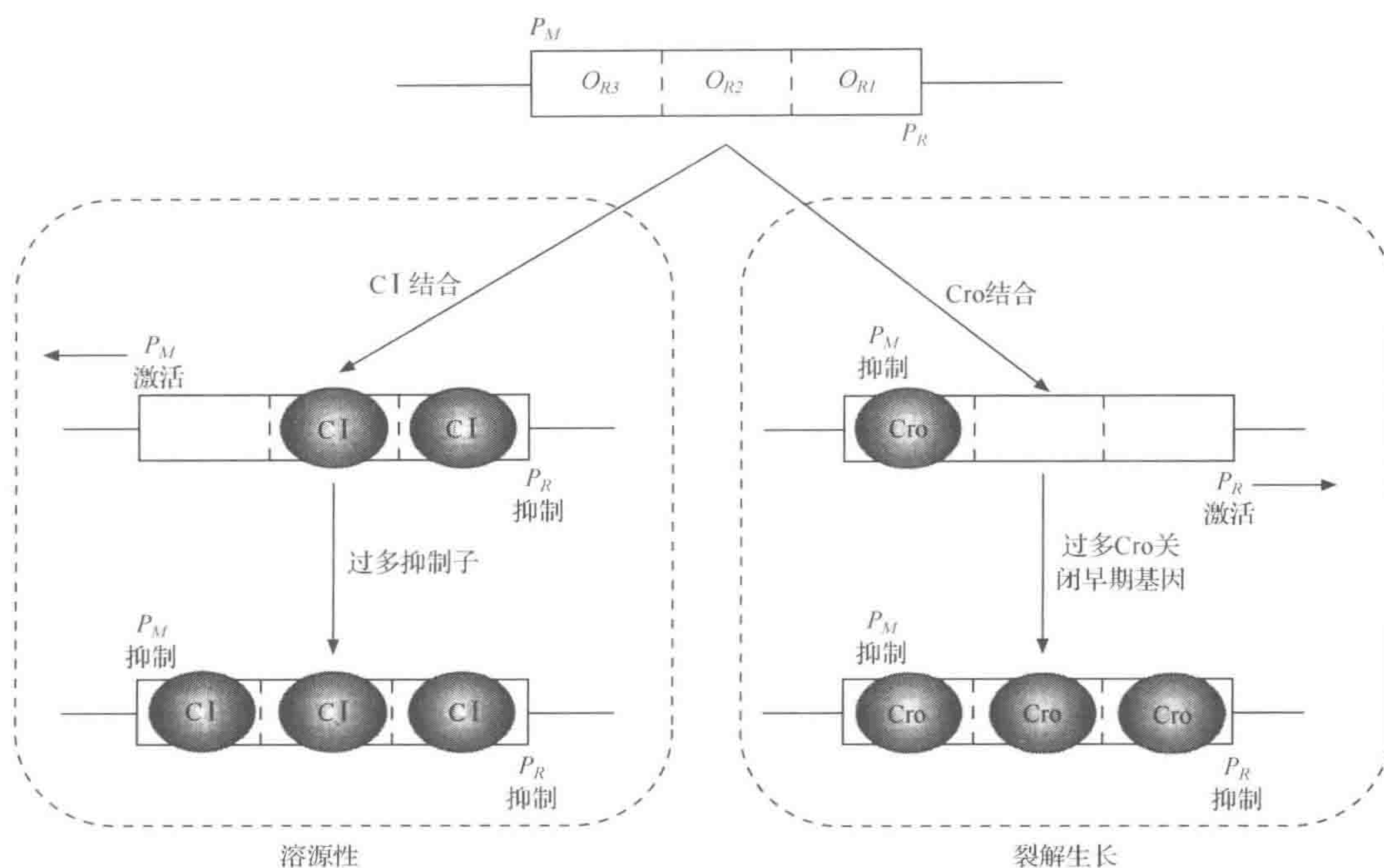


图 4.16 Cro 和 CⅠ与操纵子位点 O_R 的相互作用

由 P_E 启动的阻遏蛋白的合成。因此,增加 CⅡ稳定性和降低宿主细胞蛋白水解酶活性(延长 CⅡ活性)的突变(hfl,高频溶源)都将引起溶源频率升高。

超感染免疫性和合子诱导

溶源细菌在被噬菌体感染并溶源后,不再被更多 λ 颗粒(或同种噬菌体)感染,这就是超感染免疫性。目前,其原因已经研究得比较清楚:如果一个细胞内已存在阻遏蛋白,那么阻遏蛋白就能立即结合在进入细胞的 DNA 的 O_L 和 O_R 位点上。结果再次侵入细胞的噬菌体 DNA 在转录开始前就被抑制了。

但以溶源菌作为供体则出现相反的情况,它通过接合的形式将 DNA 转运到其他细胞中(见第 6 章)。在接合开始后的某一阶段,前噬菌体(整合在宿主染色体上的噬菌体)可作为宿主细胞染色体的一部分被传递。但是,如果受体细胞没有处于溶源状态,细胞内就不存在阻遏蛋白。这样,被传递的前噬菌体就能逃过阻遏蛋白的抑制,在受体细胞中开始新的裂解循环,进一步导致活性受体细胞数量的急剧减少。这个现象就是合子诱导。

4.5 限制和修饰

正如本章前面介绍,人们常借助某种敏感型细菌作为指示菌,通过对形成的噬菌斑个数计数来统计噬菌体的数量,其前提是噬菌体颗粒的数量和噬菌斑的个数存在直接对应的关系,也就是说,每个噬菌体颗粒都能引起一次裂解感染。在术语中,称为感染效率

(e. o. p)等于 1。但是有时并不完全是这样，感染效率会小于 1。

如图 4.17 所示，λ 噬菌体颗粒在宿主菌 *E. coli* C 细胞中培养，获得的悬浮液称为 λ. C。如果该悬浮液的培养过程用 *E. coli* C 和另一个不同的菌株 *E. coli* K 分别作为指示菌进行检测，发现在 *E. coli* C 培养基上长出 10 000 个噬菌斑(e. o. p=1)，在 *E. coli* K 中仅有 1 个噬菌斑(e. o. p=10^{-4})。但是如果用牙签从 *E. coli* K 平板上挑起一个噬菌斑在少量的缓冲液中进行重悬浮培养，得到的噬菌体培养液(称为 λ. K)同样用两种菌株分别作为指示菌，其感染效率都是 1。这时所有的噬菌体都能在两个菌株中生长。

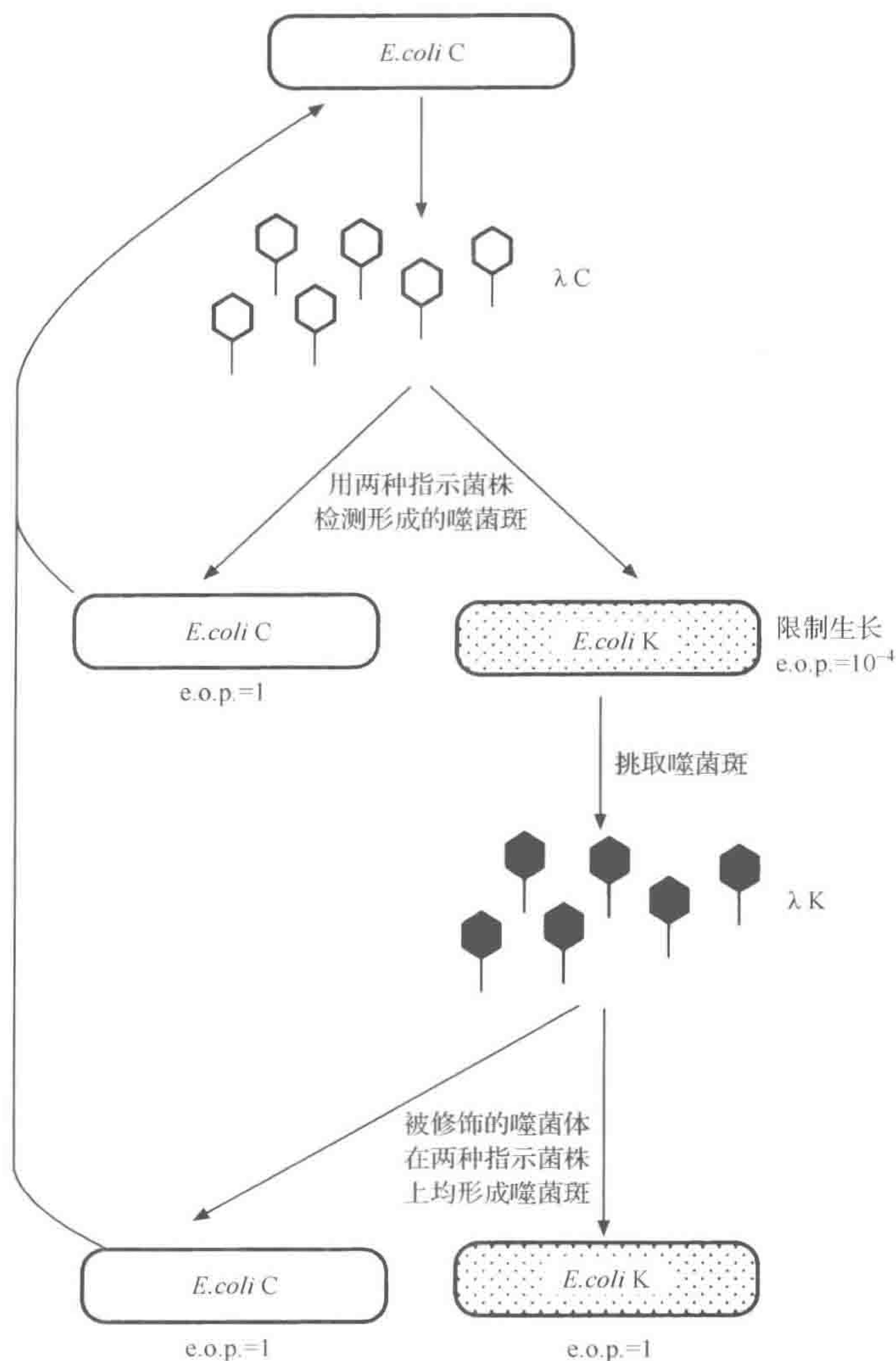


图 4.17 噬菌体宿主控制的限制和修饰作用

这种情况似乎意味着挑出一个在宿主中发生突变的噬菌体，即在初始悬液中含有少量能在 K 菌株中生长的突变体。但是，实际上用这些突变体再去感染 C 菌株，又会出现最初的情况：噬菌体在 K 菌株中生长的 e. o. p 又回复到 10^{-4}。这种行为用一般意义上的

突变几乎无法解释。

一旦了解其中的机制,所有行为的原因就很清楚了。这个现象其实就是宿主控制的限制和修饰。K 菌株产生能识别外来 DNA 的酶(内切核酸酶),并将 DNA 降解。该酶对外来细菌噬菌体的生长有限制作用,因此这种酶被称为限制性内切核酸酶。目前,人们已经知道许多这样的限制酶。大多数常用的酶属于Ⅱ型限制性内切核酸酶,它们能识别特异的 DNA 序列并在识别位点上切断 DNA 链。它们在基因操作技术中起到重要作用(见第 8 章)。*Eco*K 限制性内切核酸酶是一种Ⅰ型限制酶,它识别一串特异的序列(AAC-NNNNNNGTGC,N 代表任何一种核苷酸),但它可以在距识别位点的不同位置上切割 DNA。由此,入侵的 DNA 被限制酶降解,细胞免于噬菌体感染。

产生限制酶的宿主细胞必须有保护其自身的 DNA 不被破坏的机制。这需要另一种限制酶所识别的同样的特异碱基的酶,但不是切割而是修饰它(通常对其中一个碱基甲基化),使得这段 DNA 不再作为限制性内切核酸酶的底物。

在以上例子中,在 C 菌株(没有限制-修饰系统)上生长的噬菌体不能产生甲基化的 DNA。一旦它进入 K 菌株宿主细胞,限制性内切核酸酶将其降解;只有偶尔少数 DNA 逃过酶切作用,形成噬菌斑。噬菌斑内的病毒颗粒在修饰酶的作用下生长,DNA 受到甲基化保护,因此能感染更多的 K 菌株。然而,将它们回复感染 C 菌株后,又产生没有甲基化 DNA 的噬菌体,对 K 菌株的感染效率也降低了。

4.6 细菌对噬菌体攻击的抗性

细菌为什么能对某种特异的噬菌体有抗性,原因如下。首先,对细菌表面的受体的识别启动了噬菌体的攻击,因此,改变受体结构(或使其完全移位)的突变可能会引起抗性。其次,细菌可以通过限制和修饰系统,将噬菌体的核酸认定为“外来物”(因为噬菌体没有特异的修饰作用防止它被限制性内切核酸酶切割)。再次,温和噬菌体如 λ 能引起超感染免疫力,使得溶源菌对类似噬菌体的攻击产生抗性。

最后,在第 7 章将介绍另一个机制,由有规律地聚集在一起的一些短的散在的回文重复(缩写为 CRISPR)来完成。它们包含了插入细菌染色体上的噬菌体短片段。在转录过程中,它们转录产生与噬菌体 mRNA 互补的短 RNA 片段,而这些片段正是降解酶结合的目标片段。这也是在第 3 章中所讨论的 RNA 干扰的形式之一。

4.7 互补和重组

普遍研究的细菌的许多突变体具有影响降解或合成许多复合物的能力,而且细菌噬菌体的生长特性意味着某些噬菌体的特性可被用来作为适合的遗传标记。因为噬菌体生长需要绝大多数基因的参与,更多的目光被放在条件性突变的研究上,如使菌株对温度敏感的突变。另外一种条件性突变是在特异的基因中包含一个翻译-终止突变(如琥珀密码子)。这些突变株不能在野生型宿主菌中生长,但是可以在带有琥珀密码子抑制子的宿主菌中正常生长。

被选为非常有用的遗传标记的噬菌体的其他特性，包括噬菌斑的表型及被感染宿主菌株的范围。*E. coli* 噬菌体 T4 的 *r*Ⅱ 突变体能产生较大的噬菌斑（*r* 代表快速裂解），能在一种 *E. coli* 菌株（*E. coli* B）上生长，而不能在另一种 *E. coli* 菌株（*E. coli* K）上生长。这里不对这种菌株依赖性的作用机制进行讨论。对于这些突变体的研究，Benzer 和他的同事提供了大量相关的基因结构的信息。

T4 的 *r*Ⅱ 突变体分为两大类型：*r*Ⅱ A 和 *r*Ⅱ B。两类突变体各取一株混合后感染 *E. coli* K，形成野生型的噬菌斑（图 4.18），但是两株 *r*Ⅱ A 型突变体混合或两株 *r*Ⅱ B 型突变体混合都不能在 *E. coli* K 培养板上形成噬菌斑。这种现象不是噬菌体之间产生重组引起的，因为回收和检测子代噬菌体发现它们仍然是突变体，也就是说它们仍不能独立地在 *E. coli* K 上形成噬菌斑和在 *E. coli* B 上形成大的噬菌斑。

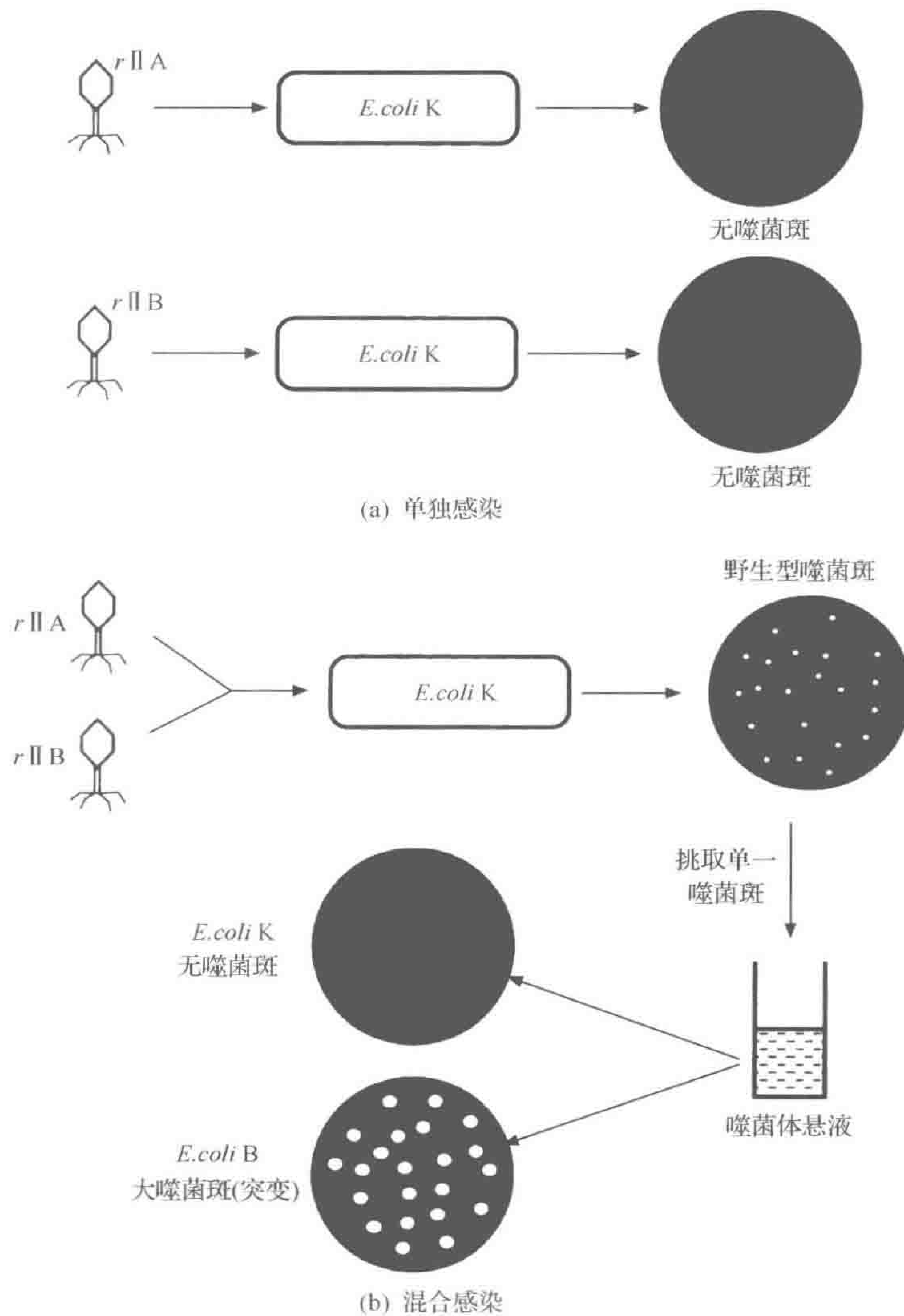


图 4.18　T4 噬菌体 *r*Ⅱ 突变株的互补作用

这个现象是互补的一个例子。*r*Ⅱ蛋白是由两条分离的多肽链组成，分别由*r*Ⅱ*A*和*r*Ⅱ*B*基因编码。混合感染时，缺失*r*ⅡA的噬菌体编码有功能的*r*ⅡB多肽链，反之亦然。这两条多肽链联合形成完整的*r*Ⅱ蛋白，因此能产生野生型的噬菌斑，但是子代噬菌体仍是突变体。互补实验完善了基因的概念，出现了顺反子这个术语。如果两个突变体有不同的顺反子，那么它们之间就会出现互补；如果它们的顺反子相同，互补就不发生。因此顺反子是指DNA上编码单个多肽链的区域。现在简单地指这段区域为一个基因，利用顺反子这个术语只是为了特别强调这一点(如多顺反子的mRNA能编码多条多肽)。在第9章中将提到更多互补的例子及其用途。

虽然互补区别于重组，但后者也会在细菌细胞受到混合噬菌体菌株感染时发生。重组率低于互补率(重组常发生在混合感染的每个细胞中)。但是，如果重组产生的子代噬菌体能在其父代所不能感染的宿主中生长的话，通过在所选的宿主细胞上培养，就能简单地检测出这些微量的重组的存在。而且，大量的噬菌体也都能被利用起来。通过这种方法，可以定位出非常临近的突变位点，有时是毗邻的核苷酸。在没有其他遗传学分析系统能达到这个水平的年代，这是一个革命性的发展。当然，现在基因克隆、DNA探针和测序等技术的结合已经使在分子水平上确定突变成为轻而易举的事。

4.8 噬菌体为何如此重要

至此，本章的重点是细菌噬菌体的分子生物学。研究噬菌体的特性不仅使分子生物学家了解细菌噬菌体本身，而且有利于分子生物学能在更广阔的领域发展。在后面的章节中，将介绍细菌噬菌体如何在基因克隆中作为载体，成为遗传学修饰的有利工具。而且，噬菌体在实验室研究、工业化生产和自然界利用等很多领域具有更多的优势。

4.8.1 噬菌体分型

所有的噬菌体的宿主范围或多或少是特定的。一般来说，它们对于某种细菌的成员(或者是亲缘关系非常近的菌种)是特异的，在某些时候这种特异性非常狭窄：它们只能感染种内的某些菌株，而不能感染其他菌株。可以利用这个特点区分同种的不同菌株，这就是噬菌体分型。

应用噬菌体分型最好的例子是区分金黄色葡萄球菌(*Staphylococcus aureus*)的不同菌株。在医院里，该细菌引起传染病的频繁发生，是一种常见的皮肤微生物，能在灰尘中存活多个星期。因此，在很短的时间内，一个病房里同时发生几起金黄葡萄球菌感染的病例可能只是个不幸的巧合。但这也可能说明在病房中存在感染源(例如，金黄葡萄球菌携带者身上脱落下大量的病原，病原又重复感染患者)。如果情况真是这样，就必须采取行动证实并控制这个源头。通过测试分离到的每种类型的菌株对某一组细菌噬菌体的敏感性，噬菌体分型提供了区分这些可能性的方法。如果传染源是单一的，感染病人的菌株就应该是相同的，就会表现出相同的噬菌体敏感性。

在诊断实验室，噬菌体也具有检测和鉴定病原细菌的功能。一般来说，如果每个感染

的细菌能产生 100 个或更多的噬菌体颗粒，那么检测的灵敏度也提高了 100 倍。而且，噬菌体的特异性也能检测特异的病原体，而非关系相近的非病原体。目前，报告噬菌体的使用增加了检测方法的敏感与方便，即利用基因工程的方法使噬菌体带上一个编码极易被检测的酶的基因。因为在有光释放的作用下极易被特异性地检测到，萤火虫的萤光素酶基因常被用作这样的基因(第 9 章将详细介绍)。该技术最大的潜能是诊断样品中较难检测的细菌，如结核分枝杆菌 (*M. tuberculosis*)，它需要数周的培养时间才形成菌落，但仍在常规检测中被广泛使用。

4.8.2 噬菌体治疗

细菌对抗生素产生抗性的问题不断增加，驱使人们寻求治疗细菌感染的替代方法。其中之一就是利用噬菌体杀死入侵的细菌。自噬菌体被发现以来，这个方法在进行不断的尝试。尽管这个想法非常吸引人，但它也面临许多问题。主要的困难是噬菌体自身是抗原，会被宿主的自身免疫系统清除(更糟的是会造成对宿主不利的免疫反应)。而且，噬菌体的宿主菌株常是特异的，如前文提到，最初敏感型的细菌会对噬菌体产生抗性，因此这个方法是否成功还不确定。那些认为噬菌体是杀死细菌的"天然"方式，细菌对其不产生抗性的说法是错误的。

然而，还有更多巧妙的方法可能达到诊治的效果。细菌噬菌体经常借助酶对细菌细胞壁的攻击来裂解细菌，那么就能用酶本身而不是整个噬菌体作为抗菌物。例如，利用噬菌体中分离的裂解酶作用于炭疽杆菌(*Bacillus anthracis*)(炭疽病的致病病原体)，治愈了实验中大部分被感染的小鼠。

4.8.3 噬菌体展示

许多蛋白质都有一个重要的特征，就是能与其他分子(配基)特异性地结合。这不仅包括酶与底物的结合，而且包括激素-受体的相互作用、抗原-抗体的结合和药物-靶标的识别。过去，蛋白质与特异配基结合是一个工作量非常大的过程。现在噬菌体提供了一种替代的方法——噬菌体展示技术。

噬菌体展示的原理如图 4.19 所示，DNA 随机片段整合到噬菌体的基因组中(常用的是丝状噬菌体 M13)，使每个 DNA 片段编码的外源蛋白质与噬菌体融合表达，并展示于噬菌体颗粒表面，最后形成噬菌体混合物(库)，在它们的表面展示不同的蛋白质。在这个混合物中，少数噬菌体表面可能展示一些感兴趣的能结合配基的蛋白质。如果用所选配基包被微孔板孔壁，所需的噬菌体能与配基结合，而其他游离噬菌体被洗去。表面富集了目标蛋白的噬菌体颗粒被洗脱下来，通过感染合适的大肠杆菌菌株得到繁殖扩增，这个过程称为淘选。通过多次重复的筛选得到纯的噬菌体制剂，提供了获得基因的来源。

利用这个技术，仅因为特异目标分子与对应多肽的亲和作用，就能从大量的噬菌体库中富集目标基因，对应多肽可以是抗原、抗体、细胞表面受体、激素或是药物。该技术对于蛋白质-配基的相互作用和药物开发的研究是一次革命。

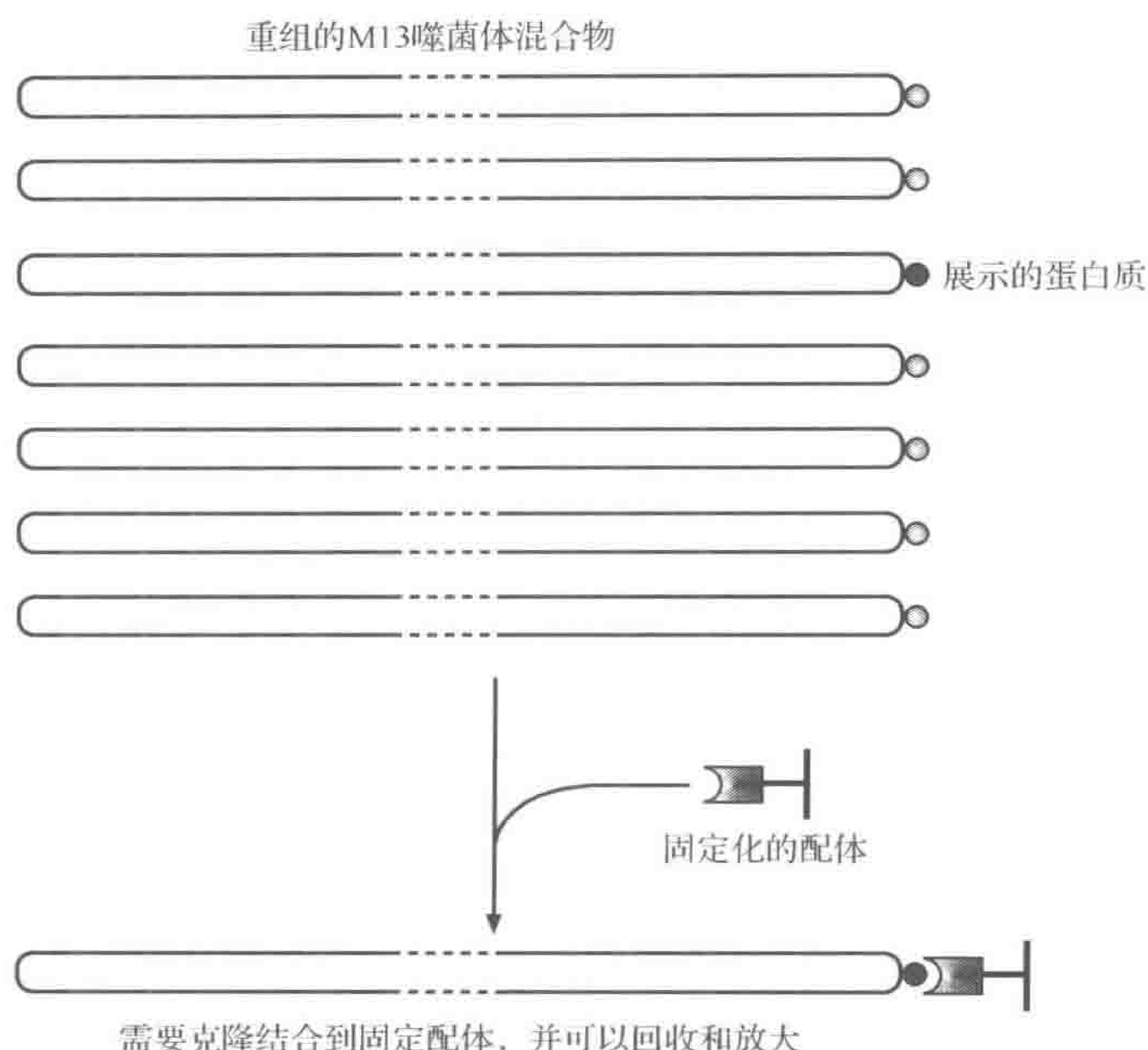


图 4.19 噬菌体展示

利用基因克隆技术将 DNA 片段插入 M13 噬菌体 DNA，产生混合的重组噬菌体，其中的一些噬菌体在其表面表达所需的重组蛋白质，这些蛋白质与衣壳蛋白以融合蛋白的形式表达。目标噬菌体表达的重组蛋白可与某一固定化的配体结合，从而将该噬菌体从混合物中分离回收

4.8.4 自然环境中的噬菌体

近百年来，噬菌体已经在实验室中得到广泛研究，但是从 20 世纪 80 年代开始，对自然环境中的噬菌体的研究才变得重要起来。最令人吃惊的发现之一是自然环境中存在大量的噬菌体。在海洋样本中，噬菌体密度是 10^6～10^7 个/L，采用类似陆生噬菌体群体的测定方法，结合海洋面积，统计得到海洋中噬菌体群体大概有 10^{31} 个颗粒。原核病毒是目前所知在地球上数量最多的生物体。这个数据说明它们的重要作用超过预期。

首先，实验研究结果表明，全球每秒钟有 10^{23} 个细菌被噬菌体感染，因此噬菌体就必须担起控制细菌群体的重任。其次，噬菌体基因组包括大量的新基因序列，所以噬菌体是一个最大的未开发的新基因库，而非其宿主细菌。最后，在后期基因转移中，噬菌体的功能不容忽视。大量的噬菌体存活在地球上意味着实验室里很少能检测到的噬菌体参与的活动在自然界中经常发生。例如，在实验室最适生长条件下，宿主细菌的 DNA 转入到噬菌体颗粒中(见 6.3 转导部分)的概率是每 10^8 次噬菌体感染出现一次。如果把全球的海洋噬菌体群体都考虑进来，就意味着在海洋中大约每秒钟发生 2×10^{16} 次个体间的基因转移。虽然在实验室以外基因转移率较低，自然界的实际转导效率也较低，但是即使在环境中只有极少部分的 DNA 实现转移，噬菌体也仍是基因转移的主要载体。

4.8.5 细菌毒力和噬菌体转化

噬菌体也有有利于细菌生长的特性。研究发现,铜绿假单胞菌中有两类基因,它们由噬菌体尾部编码,已进化为有功能的细菌素,对与其竞争的细菌有毒。有缺陷的噬菌体残余部分同样也可以促进基因转移。例如,缺陷枯草芽孢杆菌前噬菌体 PBSX,虽然大部分的自身复制功能丧失,但是仍能形成能将长约 13kb 的细菌 DNA 随机片段组装进去的噬菌体头部。与其类似,荚膜红细菌(*Rhodobacter capsulatus*)中残余的前噬菌体部分作为基因转移体,也能转移长度为 4.5kb 的细菌 DNA 的随机片段。

细菌噬菌体的另一个重要特征是它们与多种医学上重要细菌的致病性决定簇相关。例如,猩红热是由化脓性链球菌(*Streptococcus pyogenes*)的某种菌株感染引起,常伴随喉咙感染。这些症状是由细菌产生的蛋白毒素造成的。但是,这些编码毒素的基因不是细菌基因而是噬菌体基因。只有当化脓性链球菌噬菌体处于溶菌态时,才能产生毒素,因而引起猩红热。在引起白喉的白喉杆菌(*Corynebacterium diphtheriae*)中也观察到了类似的现象。毒素引发疾病的发生,特异的细菌噬菌体携带编码毒素的基因。

通过以上两个例子得知,无致病性的菌株被特定的噬菌体感染后具有致病性,这个过程称为噬菌体转化。另一个非常有意思的例子是霍乱弧菌(*Vibrio cholerae*)的转化。编码霍乱毒素亚基(*ctxA* 和 *ctxB*)的基因由丝状噬菌体 CTXφ(CTXphi)携带,而这个噬菌体与 M13 相关联(见前文)。在产毒菌株中,CTXφ 整合在染色体的特异位点上,这段位点上包含一组与毒力相关的基因。类似这样的毒力基因群在病原细菌中并不常见,把它称为毒力岛(见第 9 章)。在这个例子中,霍乱弧菌的毒力岛(VPI)其实还整合了其他的噬菌体 VPIφ,CTXφ 侵染的受体是 VPIφ 噬菌体编码的鞭毛。因此,早期受 VPIφ 的侵染是 CTXφ 侵染的必要条件。可以建立霍乱弧菌毒力变化模型(图 4.20),从一个无毒力的个体开始,整合 VPIφ 后菌株变得对 CTXφ 侵染敏感,由此导致宿主细菌获得引起霍乱的毒力基因。

细菌基因组测序揭示了先前并未了解的许多细菌噬菌体的存在,它们大多数整合在宿主的染色体上。在许多情况下,并不知道它们是否会影响细菌的表型。事实上,作者猜想大多数细菌噬菌体根本就没有这样的作用。一个典型的例外是危害性较强的借助食物传播的病原菌 *E. coli* 0157∶H7。这个菌株的基因组包含了一段附加的长 1.4Mb 的 O157∶H7 特异 DNA 序列,而在实验室温和型菌株 *E. coli* K-12 中没有。现在,已经知道这段附加的序列包括 24 个前噬菌体和类似前噬菌体的因子,它们几乎占了该序列的一半。很显然,在这个例子中,在 O157∶H7 菌株作为严重病原体出现的过程中,细菌噬菌体扮演了重要的角色。

因此,细菌的遗传学组成并不仅仅包括了染色体,在染色体序列上发生微小的变化就能形成突变体。很多细菌种类的染色体中也插入了许多噬菌体,或多或少地影响着宿主细菌的特性。下一章将介绍细菌变化的另一个主要来源,附加的、独立的 DNA 分子——质粒。在第 6 章中,将关注噬菌体在细菌间基因传递的过程中的作用。

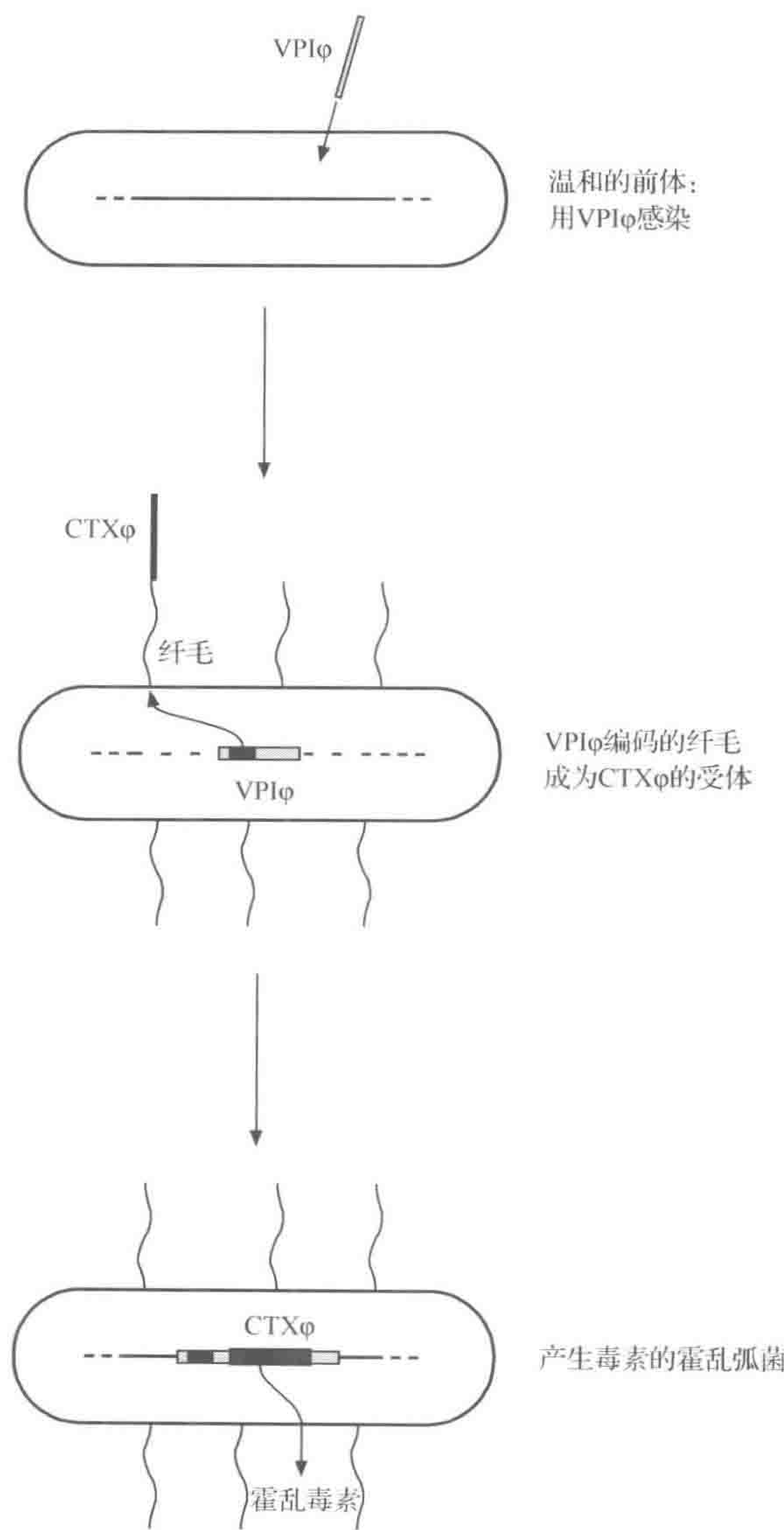


图 4.20 噬菌体和霍乱弧菌的毒力

据推测,非病源性前体被 VPIφ 噬菌体感染,VPIφ 编码的菌毛为带有毒力基因的 CTXφ 噬菌体提供了黏附位点。CTXφ DNA 插入 VPIφ 前噬菌体 DNA,产生霍乱弧菌的毒力株

5 质　　粒

在前一章中，一个细菌细胞的全部遗传组成包括了整合到染色体上的噬菌体。在对于细胞表型的影响方面，染色体外的 DNA 元件——质粒更为重要。尽管现在将其作为一个和噬菌体不同的单独现象加以考虑，但二者之间的界限并不绝对。有些噬菌体（如 P1）不整合到染色体上，但作为单独的 DNA 分子以原噬菌体状态存在，基本上就是质粒。相反，一些质粒可以相当有效地整合到染色体上，这将会在第 6 章中进行描述。

质粒和噬菌体为生物体对环境变化所做出反应的灵活性提供了一个额外维度，即环境的这些变化是敌对的（如存在抗生素）还是潜在有利的（有一种新的底物可以利用）。因此，这个额外维度包括了存在于细胞基本结构的次要的复制和生产特点。因为它们在菌株之间，或不同菌种之间的转移相对容易，从而使其在形成这些额外特性方面的作用显得尤为重要（见第 6 章）。

这是一个细菌细胞中的遗传物质组织的标准观点：它有一个单一的、环状的、携带了所有必需基因的染色体，它可能有几个比较小的染色体外的基因（质粒或噬菌体），也是环状的，携带了非必需的、在某些情况下非常有用的额外表型的基因。然而，这种观点的各个方面并非普遍真实。在某些情况下，质粒代表了全部基因组的一个相当大的比例，要么是因为有大量的不同质粒，要么质粒非常大（几百万个碱基）。质粒可能携带细胞正常生长所必需的基因。此外，还有许多菌种的质粒和（或）染色体是线性的，而不是圆形的。

莱姆病螺旋体（*Borrelia burgdorferi*）为支持以上观点提供了最好的例子，这种细菌的基因组测序结果显示，它包含了 21 个质粒，有些是圆形的，有些是线性的，它们一起组成了约 40%的细菌基因组。“染色体”长 910kb，相对较小，是线性的。在已知的所有测试条件下，至少有一个质粒是必需的，而另一些虽然不是在实验室条件下必需的，却被认为在自然条件下是必需的。也许这样一个生物体，应被视为一个分段的基因组，而不是一个单一的染色体加上许多质粒。

5.1　质粒所决定的一些细菌特性

5.1.1　抗生素抗性

最广泛研究的质粒特性是耐药性。虽然细菌对某些药物（如萘啶酸和利福平）的抗性源于编码目标蛋白的基因突变，而非源于获得的一个质粒，但还是有相当多的细菌通过获得质粒具有了抗生素抗性。质粒从一个细菌转移到另一个，有时甚至是在非常不同的菌种之间转移的能力（见第 6 章），在很大程度上造成了抗生素耐药基因的广泛传播。通过获得几个独立的质粒，或者通过获得一个具有许多抗性决定簇的单一质粒，细菌可以对多

种不同的抗生素产生抗性。在本章后面将讨论到一些例子。转座子(见第7章)通过促进基因在不同质粒间移动的方式,或从一种具有天然抗性的染色体转移到质粒的方式,在耐药质粒的发展中发挥了重大的作用。

应该知道,也会存在其他抗生素耐药性的机制,而其抗性则并非源于质粒。事实上,目前引起医院交叉感染问题的许多细菌,或者是固有的,或者是由于其对于染色体基因的抗性。

5.1.2 大肠杆菌素和细菌素

一些已被广泛研究的质粒所具有的另一个属性是能够产生一种具有抗菌作用的蛋白质,这种蛋白质通常只针对密切相关的生物。有些大肠杆菌(*Escherichia coli*)菌株产生的这样一组蛋白质,能够杀死其他的大肠杆菌菌株,因此简称为大肠杆菌素(这些概念比普通的概念,如细菌素和产细菌素,更容易为人们所熟悉,因此将用于本章)。携带在一个质粒上的大肠杆菌素基因和赋予大肠杆菌素作用免疫力的第二个基因一起,可以保护细胞抵抗自身产物的致死作用。由于对其复制和控制的详细了解,以及大多数常用的大肠杆菌克隆载体都是基于 ColE1 或其近亲构建的,一个特别的 Col 质粒——ColE1 尤其重要。

5.1.3 毒力决定簇

在前面的章节中,了解了噬菌体如何能够携带编码毒素的基因,以及噬菌体的存在对于致病性的必要性。在某些菌种中,毒力基因由质粒携带,而非噬菌体。例如,大肠杆菌的某些菌株能够引起类似霍乱的疾病(尽管比较温和)。这些菌株产生的毒素称为不稳定毒素(LT,将其与另一种热稳定的毒素区分开来)。LT 毒素与霍乱毒素密切相关,霍乱弧菌的基因由原噬菌体携带,大肠杆菌 LT 基因由质粒携带。

质粒还可以携带其他类型的毒力所必需的基因。其中最引人注目的例子之一是耶尔森氏菌属[包括鼠疫耶尔森氏菌(*Y. pestis*),鼠疫的病原微生物]一个 70kb 的毒力质粒,由于其编码一个整合系统,该系统可以使这些细菌将效应蛋白注入免疫反应细胞,以解除其武装,破坏其通信,甚至杀死它们,因此已经被形象地描述成一个移动的武器库。

另一个例子是导致炭疽病的炭疽杆菌(*Bacillus anthracis*)。其毒力依赖于两个大质粒,一个是携带编码炭疽毒素的基因的 pXO1,另一个是携带了为保护细菌抵抗宿主免疫反应的荚膜的产生和能够引起败血症所需要的基因的 pXO2。pXO2 丢失导致毒力丧失,被用做牛的一种弱毒活疫苗的 Stern 株携带 pXO1,但不携带 pXO2。相反,非炭疽芽孢杆菌菌株[如蜡样芽孢杆菌(*Bacillus cereus*)菌株和苏云金芽孢杆菌(*Bacillus thuringiensis*)菌株]获得这些毒力质粒,会使菌株只导致轻微食源性疾病或只引起昆虫炭疽病类疾病的昆虫病原体的改变。

框 5.1 提供各种致病细菌的噬菌体和质粒所携带的毒力因子的例子。但这并非是一个详尽的清单。

框 5.1　质粒和可携带毒力基因噬菌体

质粒最被熟知的特性是对抗生素的抗性。质粒(或噬菌体)也可决定其他许多特性,其中对毒力的影响是最有意义的。

下表中选择列出了一些已被较好了解的质粒和噬菌体所携带的毒力基因。但有些毒力并无明示证据是由质粒和噬菌体所决定。在某些菌株中,毒力基因可能位于染色体上。

细菌种类	疾病	毒力基因	定位
白喉杆菌	白喉	毒素	噬菌体
化脓链球菌	猩红热	毒素	噬菌体
霍乱弧菌	霍乱	毒素	噬菌体
志贺菌属某些种	痢疾	入侵/黏附	质粒
小肠结肠炎耶尔森菌	肠胃炎	Yops(外膜蛋白)	质粒
肉毒梭状芽孢杆菌	肉毒中毒	毒素	噬菌体
破伤风梭状芽孢杆菌	破伤风	毒素	质粒
肠埃希氏菌(大肠杆菌)	肠胃炎	肠毒素	某些质粒
肠埃希氏菌(大肠杆菌)	肠胃炎	黏附	某些质粒
炭疽芽孢杆菌	炭疽	毒素,荚膜	某些质粒

5.1.4　植物相关细菌的质粒

致病性的一个不同类型是导致肿瘤生长(在一些植物被称为冠瘿)的植物病原——根癌农杆菌(*Agrobacterium tumefaciens*)。而且只有携带特定类型质粒(已知的导致肿瘤的 Ti 质粒)的菌株才具有致病性。而在这种情况下,致病性与质粒自身 DNA 的一个特定部分转入植物细胞相关。因其在植物细胞遗传操作领域的应用,这种现象尤其重要。

尽管是共生而非致病,根瘤菌属成员可以"感染"植物。这些细菌在豆科植物的根部形成结节。在这些条件下,这种细菌能固氮并为植物提供一种还原态氮的有用资源,而这对于生态和农业都相当的重要。质粒携带了为结瘤和固氮所必需的基因。

5.1.5　代谢活性

质粒能够以其他各种方式扩大宿主细胞代谢活动的范围。例如,如果将一个携带发酵乳糖的基因的质粒引入到非发酵乳糖菌株,将会使其转换成为一个能够利用乳糖的菌株。在以有限的生物化学特点作为鉴定基础的诊断实验室,这样的质粒会引起一些麻烦。由于沙门氏菌发酵乳糖的不稳定性,最初常用于把潜在的致病性沙门氏菌属从非致病性大肠杆菌中鉴别出来。在某些情况下,由于病原体获得了乳糖发酵质粒,沙门氏菌感染的严重流行病的检测已被推迟。

在质粒上也发现了其他的大量基因,包括发酵其他糖类(如蔗糖),水解尿素或产生硫化氢。其中许多最初得以鉴定只是因其在生化鉴定测试中引起了混乱。

生物降解和生物修复

质粒介导的代谢活性的另一种类型是降解潜在有毒化学品的能力。一个来自恶臭假单胞菌(*Pseudomonas putida*)的这种类型的质粒 pWWO,编码将环烃甲苯和二甲苯转换为苯(甲)酸盐(上游途径)的一系列酶,并且通过儿茶酚中间体的环裂解,负责将苯甲酸降解为能够用于产生能量和生物合成的可代谢中间体的第二个操纵子(下游途径)(图 5.1)。因此,这种微生物可以以甲苯作为唯一碳源生长。上游途径的酶是特异的,其他质粒编码上游途径具有特异性的酶,使微生物能够将其他化学品转化为可以被下游途径的酶降解的苯甲酸和儿茶酚衍生物。质粒介导的降解包括萘、樟脑以及氯代芳香族化合物,如 3-氯苯和 2,4-D 除草剂 (2,4-二氯苯氧基乙酸)。

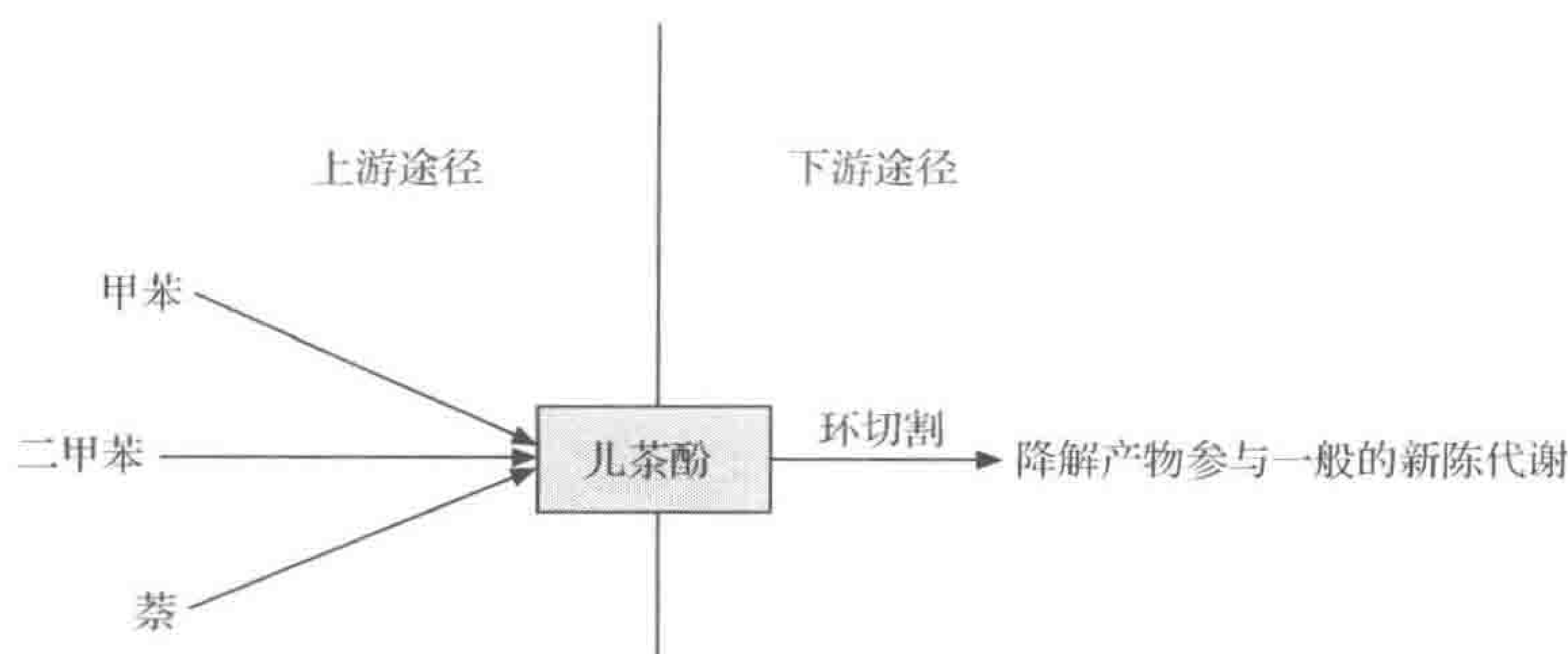


图 5.1 环状碳氢化合物的降解

降解破坏环境的化学物质的能力在清理污染场地方面具有潜在的应用价值(生物修复)。扩大微生物所降解的化学物质的范围,通过修改现有途径,从受污染的地点筛选出具有新活性的细菌,正越来越多引起人们的注意。由于质粒具有对有毒金属离子,尤其是铜和汞的抗性,这样的菌株也具有重要的潜在应用价值。

5.2 质粒的分子特性

由于 DNA 的超螺旋,细菌质粒一般以构象非常紧凑的环状 DNA 分子形式存在于细胞内。在分离时(尤其是大质粒)可能发现,当其中一股断裂时,超螺旋丢失,质粒松弛成为不紧凑的开放环型。在许多情况下,质粒是相当小的分子,只有数千碱基长度,但在某些微生物中,特别是假单胞菌属的成员,长达几百万碱基的质粒是常见的。然而,值得注意的是,分离质粒的标准方法是适于小的共价闭合环状 DNA,大质粒或者如线性质粒等其他形式质粒的出现,有可能被低估。

可以方便地将来自大肠杆菌的质粒分为两种类型。以 ColE1 为原型的第一种质粒比较小(通常小于 10kb),在细胞中有多个拷贝。虽然对于质粒复制有一些控制(本章后面部分将进行讨论),其复制与染色体复制和细胞分裂的过程不相关(因此拷贝数比较高)。在阻止染色体复制的一定条件下(如抑制蛋白质合成),这些质粒的复制可以继续,

从而使每个细胞内的质粒拷贝数大量增加。当要分离想要得到的质粒时，这种被称为质粒扩增的现象是非常有用的。

第二种类型以 F 质粒（F 代表育性）为例，都比较大（一般超过 30kb，F 质粒大约 100kb），通常在每个细胞中只有一个或两个拷贝。这是因为其复制和染色体复制一同要受到控制，因此当一轮染色体复制启动时，也会出现质粒的复制。许多这样的大质粒通过接合能够促进自身的转移。

基于其不同的生存策略，这两种类型质粒的存在都是合理的。第一种类型的质粒成员依靠其高拷贝数，可以确保在细胞分裂时，如果质粒分子随机分配到两个细胞，每个子细胞几乎可以肯定至少包含一个复制的质粒[图 5.2(a)]。例如，每个细胞中一个质粒如果有 50 个拷贝，一个子细胞接受不到质粒的任何拷贝的机会低达 $1/10^{15}$。

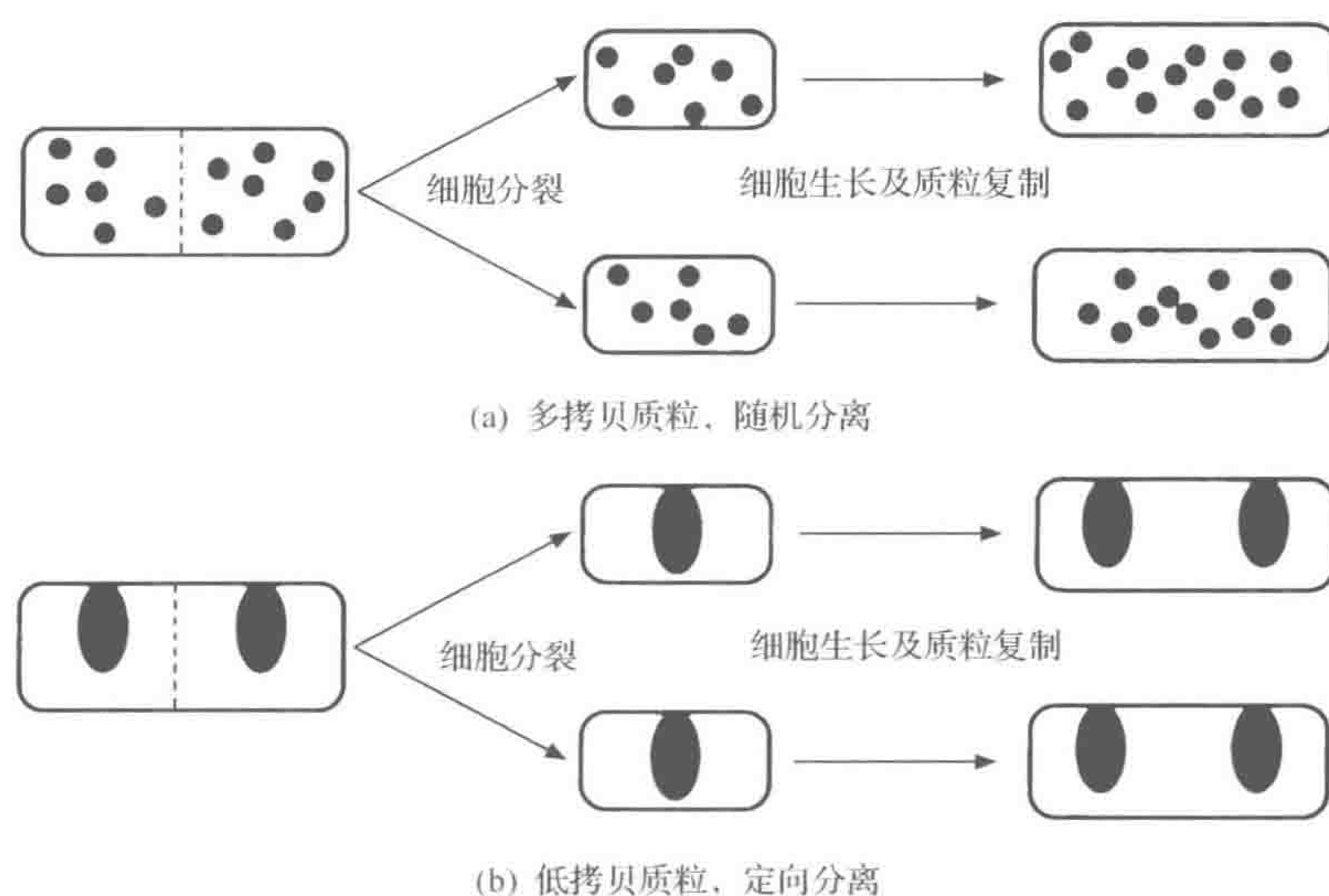


图 5.2　细胞分裂时质粒的分离

然而，高拷贝数约束了质粒大小的范围。质粒复制造成了关系到质粒大小和拷贝数的新陈代谢的负担。负担越重，那些细胞不具备质粒的选择性压力越大。因此，高拷贝数质粒比较小是合乎逻辑的。例如，ColE1 大小是 6.4kb。如果每个细胞中有 30 个拷贝，这表示其约为细胞总 DNA 的 4%。另外，F 质粒（大约 100kb），如果它在每个细胞中的拷贝数与之类似，将增加近 70%的总 DNA 含量，这将不可避免地使细胞生长速度慢很多，而且失去了质粒的细胞将具有明显的选择性优势。

对于大肠杆菌而言，建立接合的所需信息是相当广泛的（见第 6 章）。例如，在 100kb 的 F 质粒中有大约 30kb 是由质粒转移所需的基因组成。因此，一个小质粒将无法携带进行接合转移所需的所有信息。

第二种类型的质粒以不同的策略进化[图 5.2(b)]。首先，将质粒的复制和染色体的复制相关联，以确保在细胞分裂时至少有两个拷贝的质粒。其次，随机分区将不足以确保每个子细胞接收一个拷贝；因此，质粒必须以定向的方式在后代之间分布。后面会考虑如何做到这一点。因为任何源自分区机制失效群体的无质粒细胞可以充当质粒转移的接受

者，接合转移的能力提供了一种备份机制。

虽然图 5.2(b)有助于我们的认识，但即使是在大肠杆菌，这也是一张高度简化，存在许多例外的图，清楚这一点非常必要。尽管它们都没有接合，以及有一些以多个拷贝形式存在的较大的质粒的例子，但低拷贝数小质粒的例子不胜枚举。另外，其他生物的描述更不太清楚，例如，在链霉菌中，似乎相当小的质粒就能够通过接合促进其自身的转移。

5.2.1 质粒的复制和控制

如上所述，为了了解质粒的不同行为的原因，需要了解质粒的复制机制及其如何受控。这应该与在第 1 章中对于染色体复制的描述相比较。许多质粒以双链环状分子的形式复制。这些质粒复制的全面描述与染色体复制基本相似，因为复制起始于一个固定的点，*oriV*（植物起源的，将其与接合转移位点起始区分开来，*oriT*）。从这个点开始，或者向一个方向，或者同时向两个方向复制，直到整圈。然而，复制的某些方面与染色体复制并不相同。已深入研究的两个例子是 ColE1 和 R100。其他具有完全不同的复制模式的质粒将在本书后面进行讨论。

ColE1 的复制

具有产生大肠杆菌素能力的质粒 ColE1（图 5.3）是一个相对较小的质粒（6.4kb），携带了产生大肠杆菌素以及针对其免疫力的基因，共同具有维持质粒的功能。这可能是所有质粒中最好理解的例子。复制起始于从 *oriV* 上游 555bp 处的一个位点开始的 RNA 引物的产生（RNAⅡ）（图 5.4）。转录从 *oriV* 处开始，在一个特定的位点，RNaseH 切割 RNAⅡ（当其以 RNA-DNA 杂交体形式存在时，切割 RNA）。最后，通过向 RNA 引物的 3′-OH 端添加脱氧核苷酸合成 DNA。

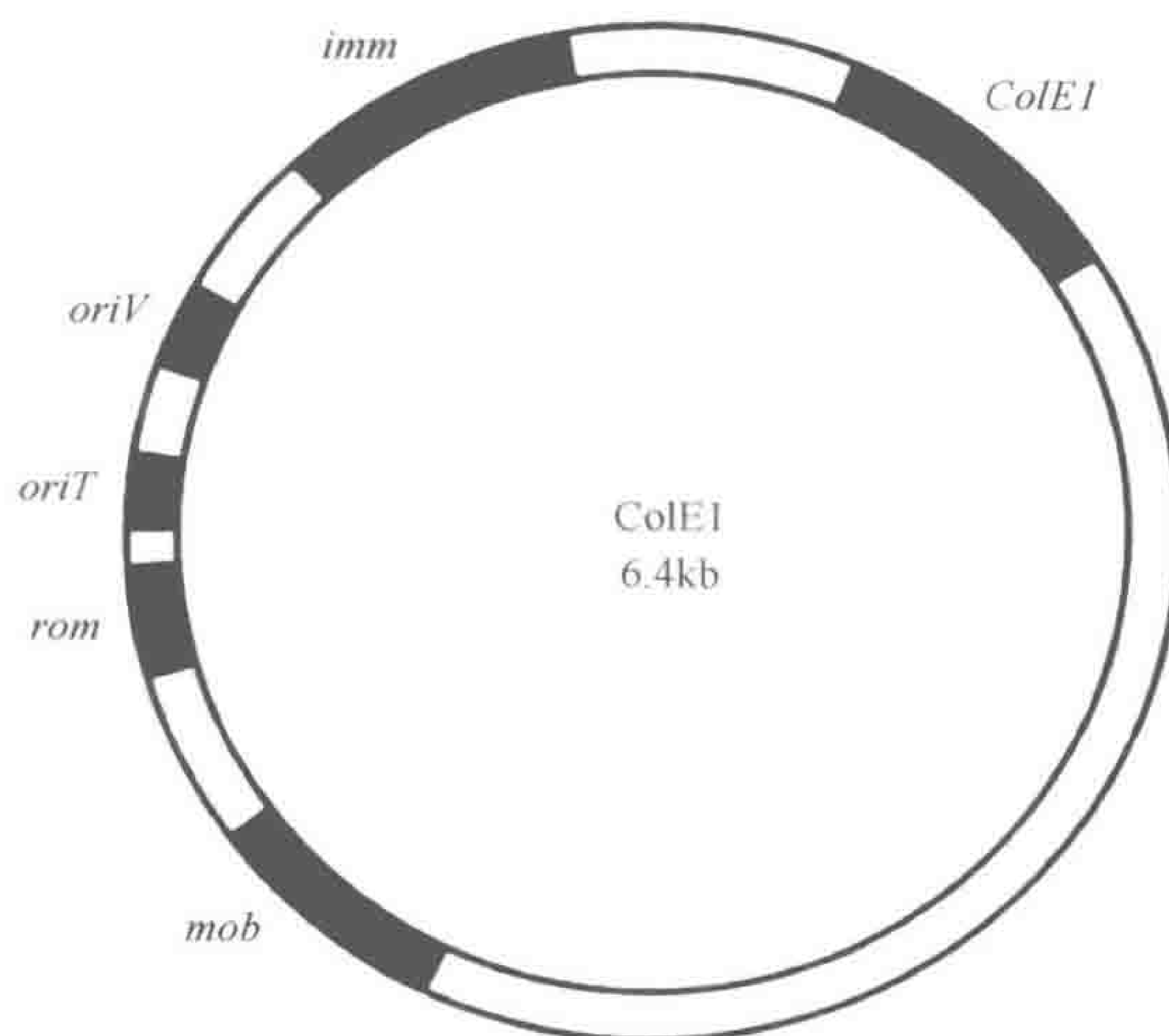


图 5.3 ColE1 质粒的遗传图

colE1 及 *imm* 基因编码大肠杆菌素 E1 及对其的免疫力；*mob* 编码质粒动员所需的核酸酶；*rom* 编码有效控制拷贝数的蛋白质；*oriT* 是接合转移的起点；*oriV* 是复制起点

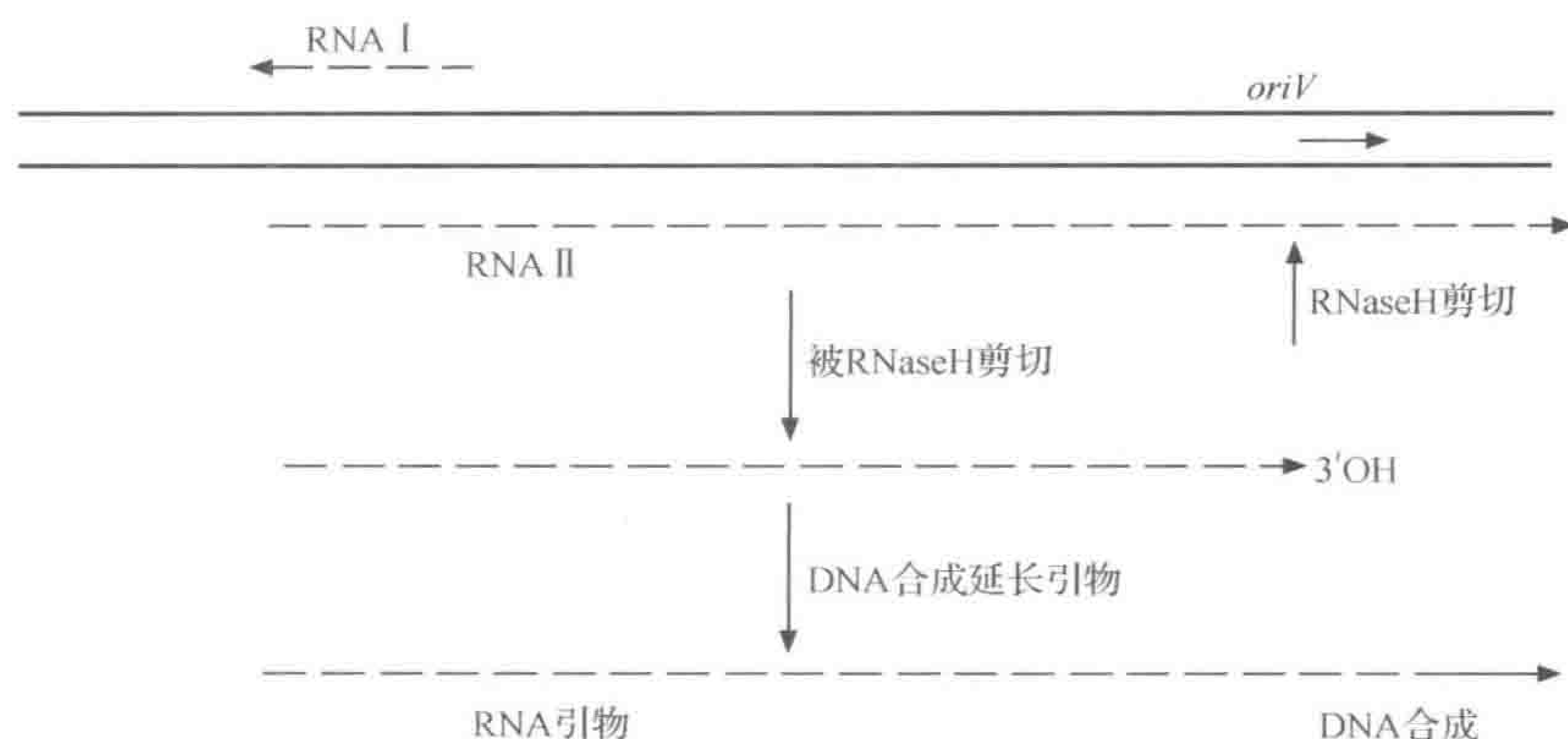


图 5.4 ColE1 质粒复制起点的结构与控制

RNA Ⅱ是 DNA 合成的引物，随后被 RNaseH 剪切。RNA Ⅰ与 RNA Ⅱ结合，抑制 RNase 的剪切，因此阻止复制的起始

有另一种 RNA 分子，即在同一区域产生的 RNA Ⅰ。这是从 RNA Ⅱ 的反向链转录，并与 RNA Ⅱ 开始的 108 个碱基互补。因为 RNA Ⅰ 结合到 RNA Ⅱ，干扰其二级结构，阻止了 RNaseH 剪切 RNA Ⅰ，从而抑制复制。另外一个控制复制的基因是编码一种促进 RNA Ⅰ 和 RNA Ⅱ 相互作用的蛋白质的 *rom* 基因。*rom* 基因被删除的 ColE1 衍生质粒的拷贝数更高。

ColE1 质粒是非接合的：其自身不能从一个细胞转移到另一个细胞。然而，和许多其他非接合质粒一样，如果细胞带有一个兼容的接合质粒，它可以通过接合转移。这涉及 *mob* 和 *oriT* 位点的效应，被称为可动化，将在第 6 章予以描述。

R100 的复制

R100(图 5.5)是低拷贝、接合型的抗性质粒，其中包含约 89kb 的 DNA，对 4 种不同的抗生素(四环素、氯霉素、链霉素、磺胺)以及汞盐具有抗性。所有为接合转移所需的基因都聚集在一起，并与复制起始位点(*oriC*)相邻，组成了近一半的质粒，而耐药性决定簇都位于图 5.5 的右半部分。这些大的耐药性质粒通常以这种方式组合起来，这反映了其通过将耐药基因按顺序添加到一个基本复制子的进化方式，也就是说，它起源于一个神秘的质粒，各种抗性基因被添加到这个只包括转移区和复制起点的质粒(单个或多个)。

获得额外抗性基因的机制之一是四环素耐药性决定簇(*tet*)，其侧翼各有一条插入序列(IS10)，这个组合形成了可以从一个 DNA 位点移动到另一个 DNA 位点被称为转座子的移动结构(在这种情况下是 *Tn10*)。R100 还包含了一条不同的插入序列(IS1)的两个拷贝。转座子和插入序列在质粒的进化中扮演了关键的角色，这将在第 7 章中做进一步的讨论。

从一个单一的起点 (*oriV*) 开始的 R100 复制，需要 *repA* 基因的产物(图 5.6)。两个基因调节着 RepA 蛋白的产生。其中之一，*copB* 编码一种抑制 *repA* 基因转录的蛋白质。当质粒首次进入细菌细胞，CopB 的缺失允许 RepA 表达，所以有一段时间的复制，直到 CopB 积累到抑制该启动子的水平。第二个调节基因，*copA* 编码 80～90 个核苷酸未翻译

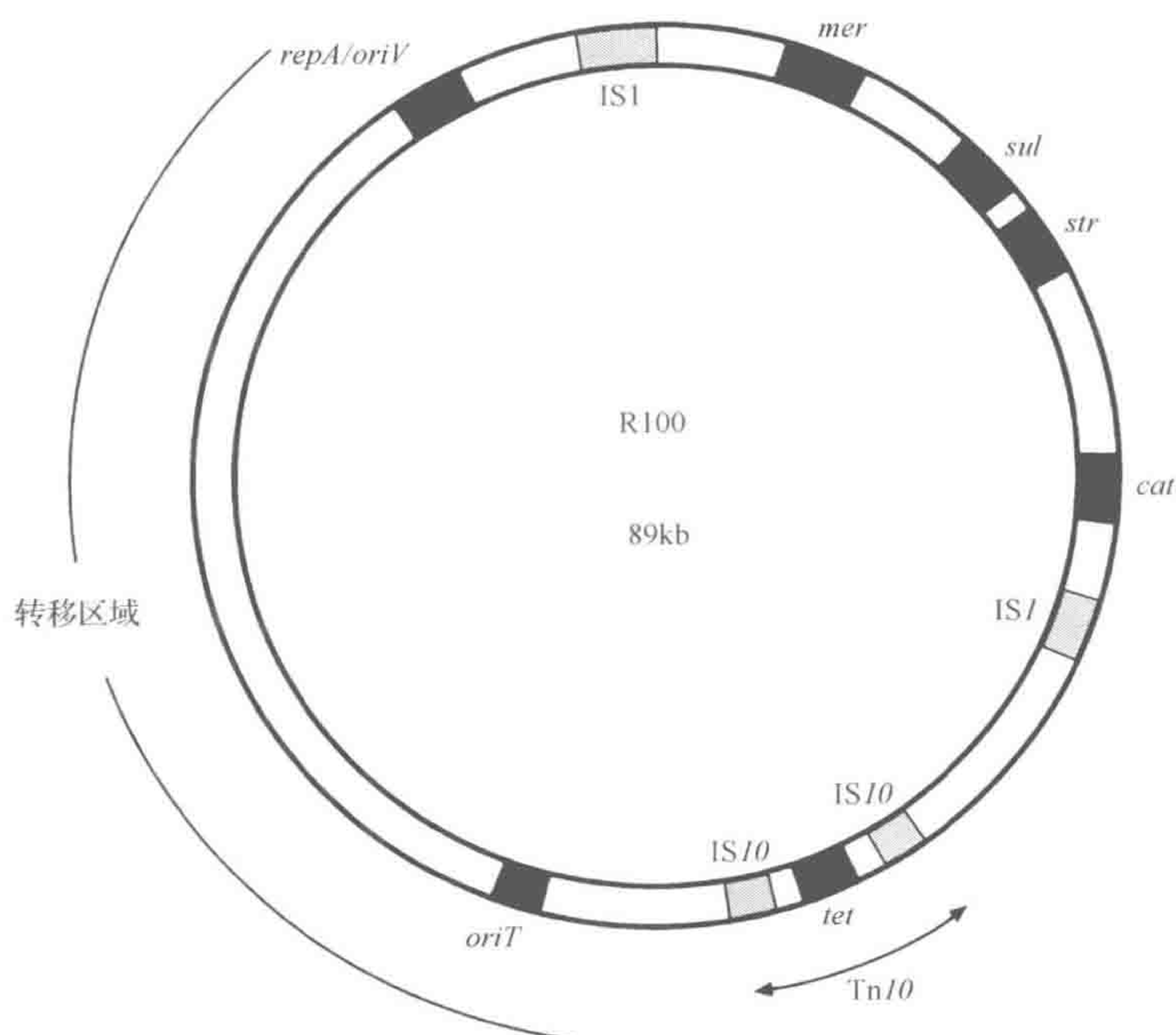


图 5.5 大肠杆菌接合质粒 R100 的遗传图

抗性基因：*cat* 为氯霉素(氯霉素转乙酰酶)；*met* 为汞离子；*str* 为链霉素；*sul* 为磺胺类药；*tet* 为四环素。其他位点：*oriT* 为接合转移起始位点；*repA*/*oriV* 为复制功能和复制起点。IS*1* 和 IS*10* 为插入序列；*Tn*10 为转座子

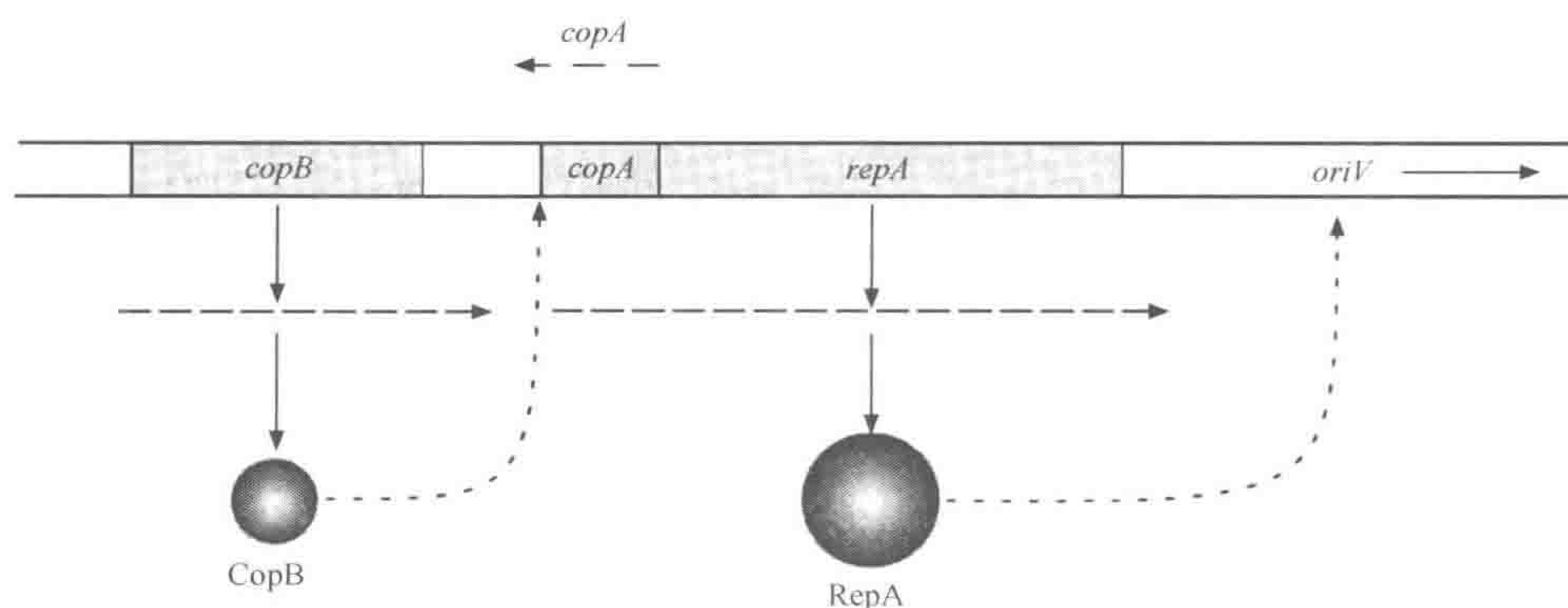


图 5.6 质粒 R100 的复制控制

复制起始时需 RepA 蛋白。CopB 抑制 *repA* 的转录，*repA* mRNA 的翻译可被反义的 *copA* RNA 抑制

的 RNA 分子。*copA* 基因位于转录产生 RepA 的 DNA 区，但以相反的方向转录(是反义 RNA)，因此，*copA* RNA 与 *repA* 转录本的短区互补，并将与其结合，干扰 RepA 的翻译。当质粒复制时，*copA* 基因的拷贝数加倍，*copA* RNA 的数量因此增加；这导致了下一步复制起始的显著降低，直到细胞分裂恢复了最初的拷贝数。

通过 DNA 重复序列控制质粒的复制(质粒复制蛋白结合位点)

如上所述，通过控制 RepA 的表达调控质粒的复制对于许多低拷贝数的质粒是足够

的。但是,质粒数量和其复制能力之间缺乏直接的联系,这就意味着预期的拷贝数将会分布在平均值的狭窄两侧。例如,大多数细胞可能有5个拷贝,但有些会有6个,有些会有4个。在一种积极而有效的分区系统存在的情况下,这对于这种质粒的稳定遗传是足够的。但有一些质粒,如F质粒,只以一个拷贝的形式存在,细胞分裂后,每个循环复制一次,使细胞分裂时有两个拷贝。没有了质粒拷贝数的统计分布的余地。这种质粒的稳定维护要求更严格的、与细胞中的拷贝数直接相关的复制控制。

这种控制是由存在于复制起始区域,被称为质粒复制蛋白结合位点(iteron)的17～22bp长的重复DNA序列所提供的。例如,F质粒有9个17bp的重复序列。RepA蛋白结合到这些iteron序列。当有一个以上的质粒时,RepA蛋白可以在两个拷贝上结合iteron,使其耦合在一起。这可以防止质粒的进一步复制(图5.7)。这种"耦合"或"手铐"模式提供了极其严格的控制质粒拷贝数的一种机制。

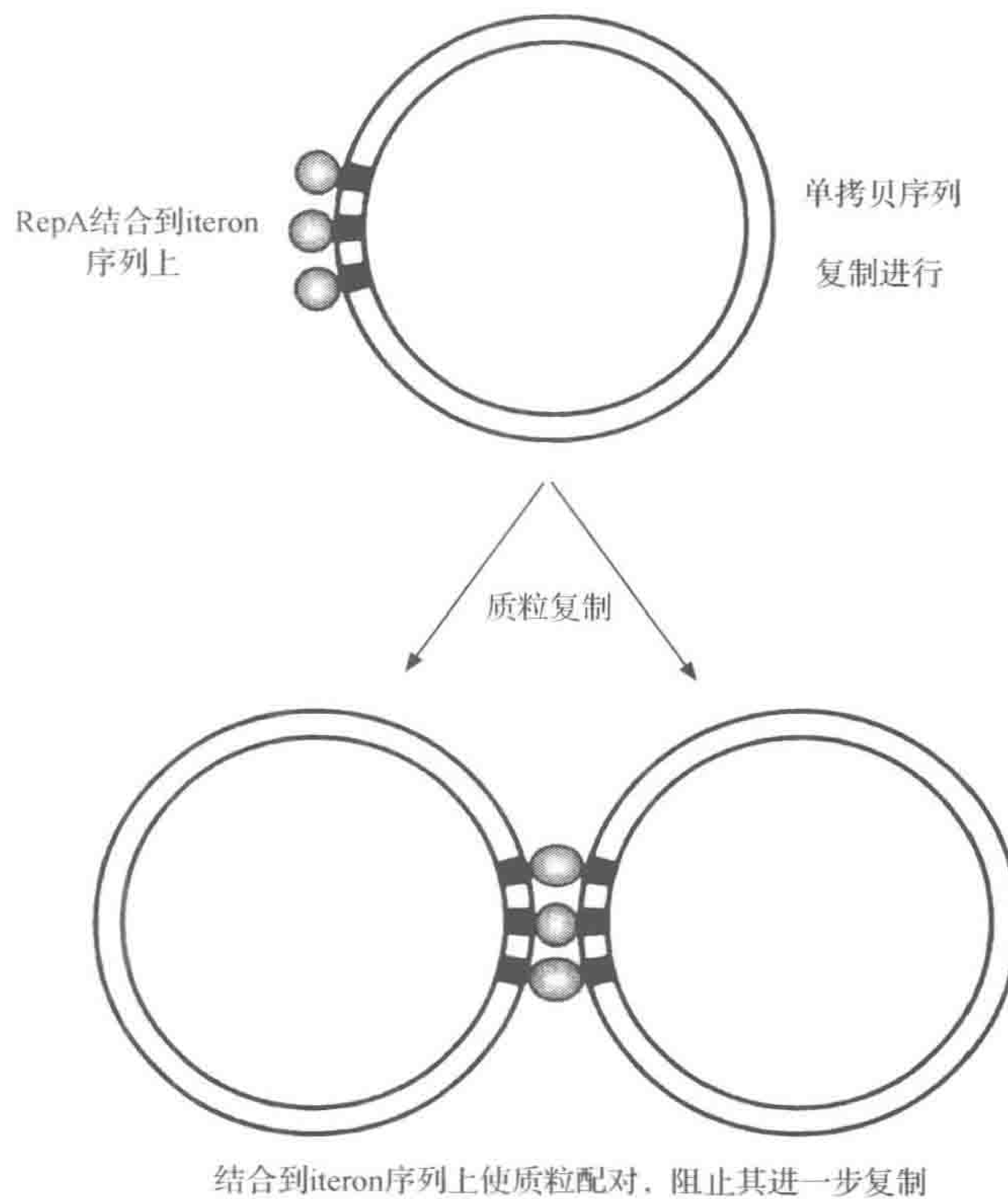


图5.7　含有iteron序列的质粒的配对控制模式

单链形式的质粒复制

大肠杆菌质粒的复制通常似乎遵循着在这一章前面部分所描述的路线,即作为同一进程中的一部分复制两条链,这种模式并非是普遍适用的。许多质粒,尤其是革兰氏阳性菌的质粒,通过单链中间体进行复制(图5.8)。尽管这些质粒通常被称为单链质粒,单链的形式只是复制中间体,而通常大部分质粒分子是双链的。

已经遇到过类似的复制模式,即单链噬菌体φX174和M13的复制模式(第4章)。

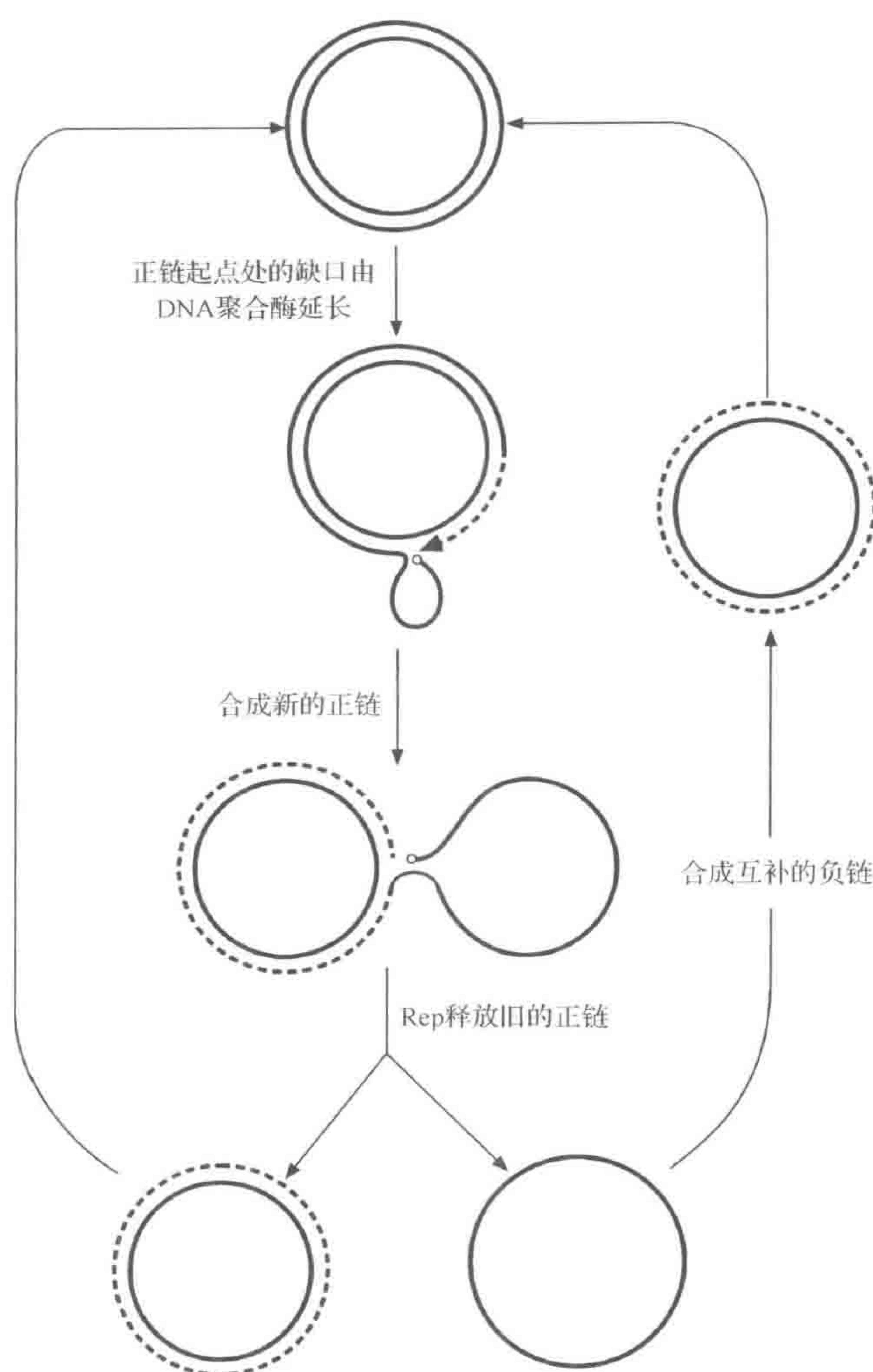


图 5.8 单链质粒的复制

不仅在一般的概念,而且还在所涉及的机制中,这些系统之间都有着很大的相似性。虽然在一些细节上有所不同,所有这些元件复制的一般机制都遵循图 5.8 的描述,质粒编码的蛋白 Rep 首先切割正链上的特定位点。有缺口的 DNA 提供了启动 DNA 合成的位点,利用宿主的酶取代了旧的正链。虽然这种情况正在发生,Rep 蛋白仍然结合到有缺口正链 5′游离端。当这一过程绕完质粒回到正向起点,Rep 蛋白制造了另一个缺口,释放旧的正链,连接这个分子的末端产生一个完整的单链环状结构。之后通过开始于另外一个起点的(负链起点)互补链(负链)的合成转换为双链形式。

同样的模式可以适用于单链噬菌体(见图 4.5)。重要的区别是,这些噬菌体的单链形式不只是一个复制中间体结合到噬菌体颗粒。在复制的早期阶段,通过被称为单链 DNA 结合蛋白(SSB)的宿主蛋白的包备,稳定了所替代的单链形式。在周期的后期,子代病毒被包装进噬菌体外膜,而不是产生更多的 RF DNA。φX174 的病毒链 DNA 被直

接包装进噬菌体的头部,互补链的合成被阻止。M13 的一个特定的噬菌体蛋白(基因 V 蛋白)积聚并且替代病毒 DNA 上的 SSB,阻止了互补链的合成。当噬菌体在细胞膜上组装时,噬菌体颗粒的蛋白质替代了一种基因 V 蛋白。

许多这些元件的正向起点的序列已经确定,可以分为彼此高度相似的几个组。例如,噬菌体 φX174 和金黄色葡萄球菌的质粒 pC194 和 pUB110 都在起点具有相同的短序列。这是这些元件的 Rep 蛋白识别位点,序列也具有相似性。这表明了噬菌体 φX174 和来自革兰氏阳性菌的质粒之间的进化联系。

另外,负向起点更加多样化。由于该起点需要宿主蛋白的作用,它必须进化以适应目前宿主的特殊性。一个后果是,如果这种质粒被置入一个不同的菌种,正向起点的作用可能发挥得相当不错,但需要宿主蛋白质的负向起点就可能会无效。这不仅会由于分离不稳定性导致质粒的频繁流失,而且由于积聚的单链 DNA 形式将刺激重组,会产生多种 DNA 重组。

线性质粒的复制

另一种类型的质粒是对细菌的 DNA 结构和复制的一般认识的巨大挑战。源自包括伯疏氏螺旋体属和链霉菌属的几个菌种具有线性 DNA 的特点。在这些细菌种类中(和一些其他的菌种),染色体也是线性的。这些线性分子的复制提出了一些问题。如前所述(第 1 章),一个新的 DNA 链(前导链)是连续产生的,而另一条链(后随链)是不连续产生的;也就是说,它以短片段的形式产生,随后连接在一起。由于核酸总以 5′→3′方向合成,所以这是必要的。DNA 的合成通常需要由 RNA 聚合酶产生的引物。当合成下一个片段时,这些序列被移除,并被到达该处的 DNA 聚合酶所取代。

当复制叉接近线性 DNA 分子的末端时,发生了什么呢? 前导链将继续向右直到末端,但后随链会如何呢? 恰好在模板链的 3′端产生 RNA 引物是不可能的,并且在任何情况下,没有用 DNA 替代 RNA 最后一个碱基的机制。因此,模板链 3′端不能被复制(图 5.9)。

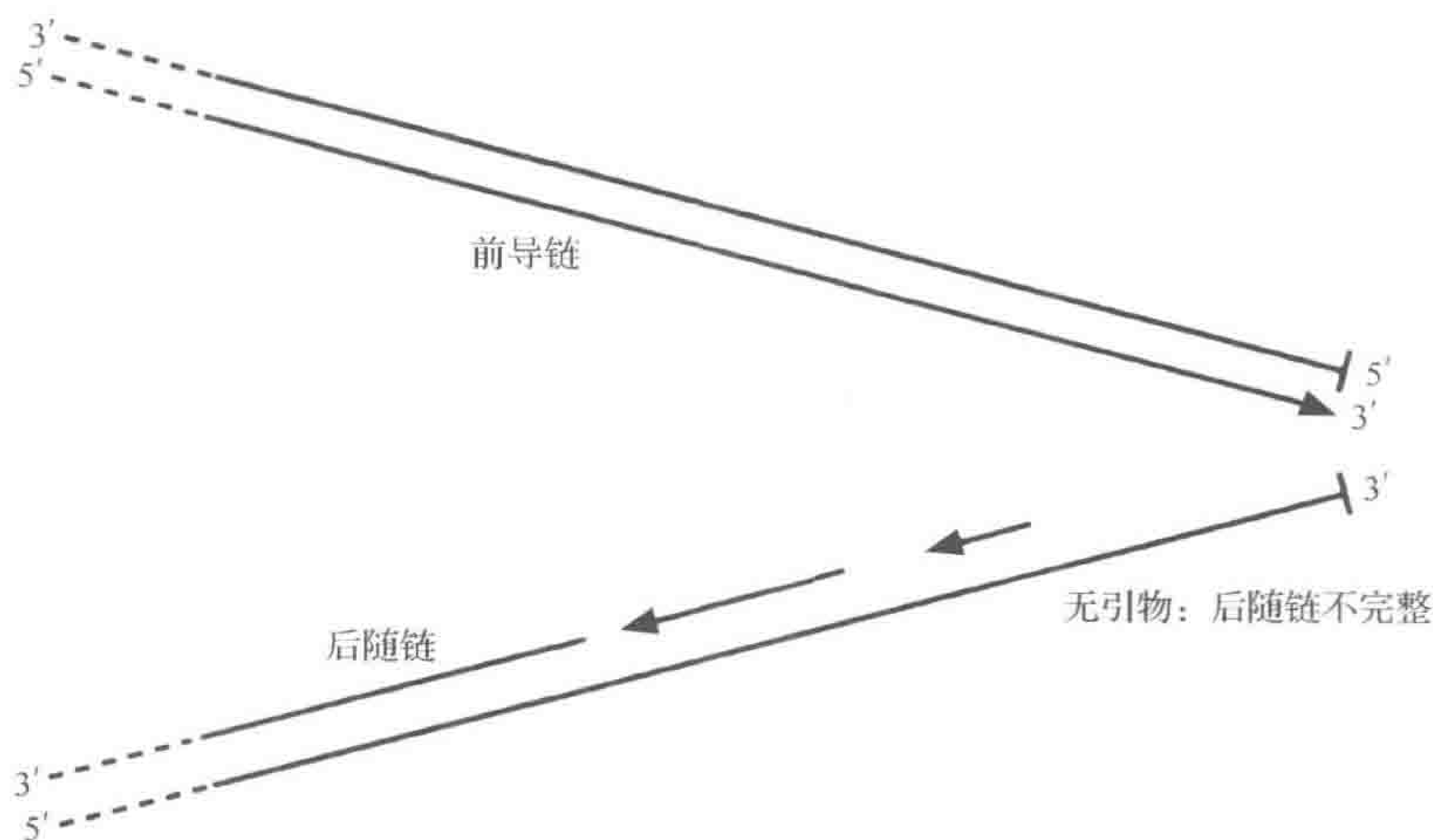


图 5.9　线性 DNA 的不完全复制

普通的复制模式不能复制出线性分子的末端

线性 DNA 的细菌(质粒或染色体)采取了不同的策略来解决这个问题。伯疏氏螺旋体属的两条链的两端以共价闭合的发夹结构连接在一起(图 5.10)。在一些动物病毒(如痘病毒)及前噬菌体是一个线性的 DNA 质粒的大肠杆菌噬菌体 N15 中,也发现了类似的结构。图 5.11 说明了这样一个结构的完整复制的一个模型。在复制的中央起点(*oriC*)启动的双向复制会导致一种二聚体双链环状分子的形成,基因组的两个拷贝由发夹环序列的拷贝连接起来。通过在每个末端切割和重新连接 DNA 链以重新形成共价闭合的发夹环,该中间体结构被加工。

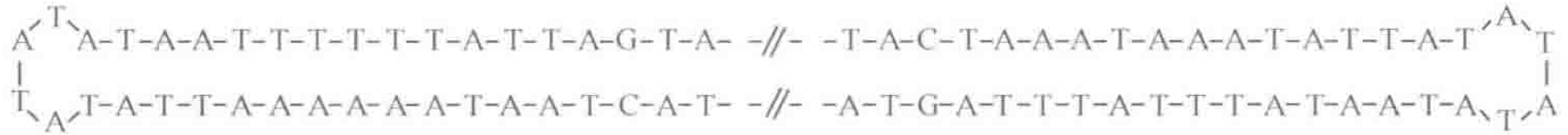


图 5.10　伯疏氏螺旋体线性质粒末端共价连接

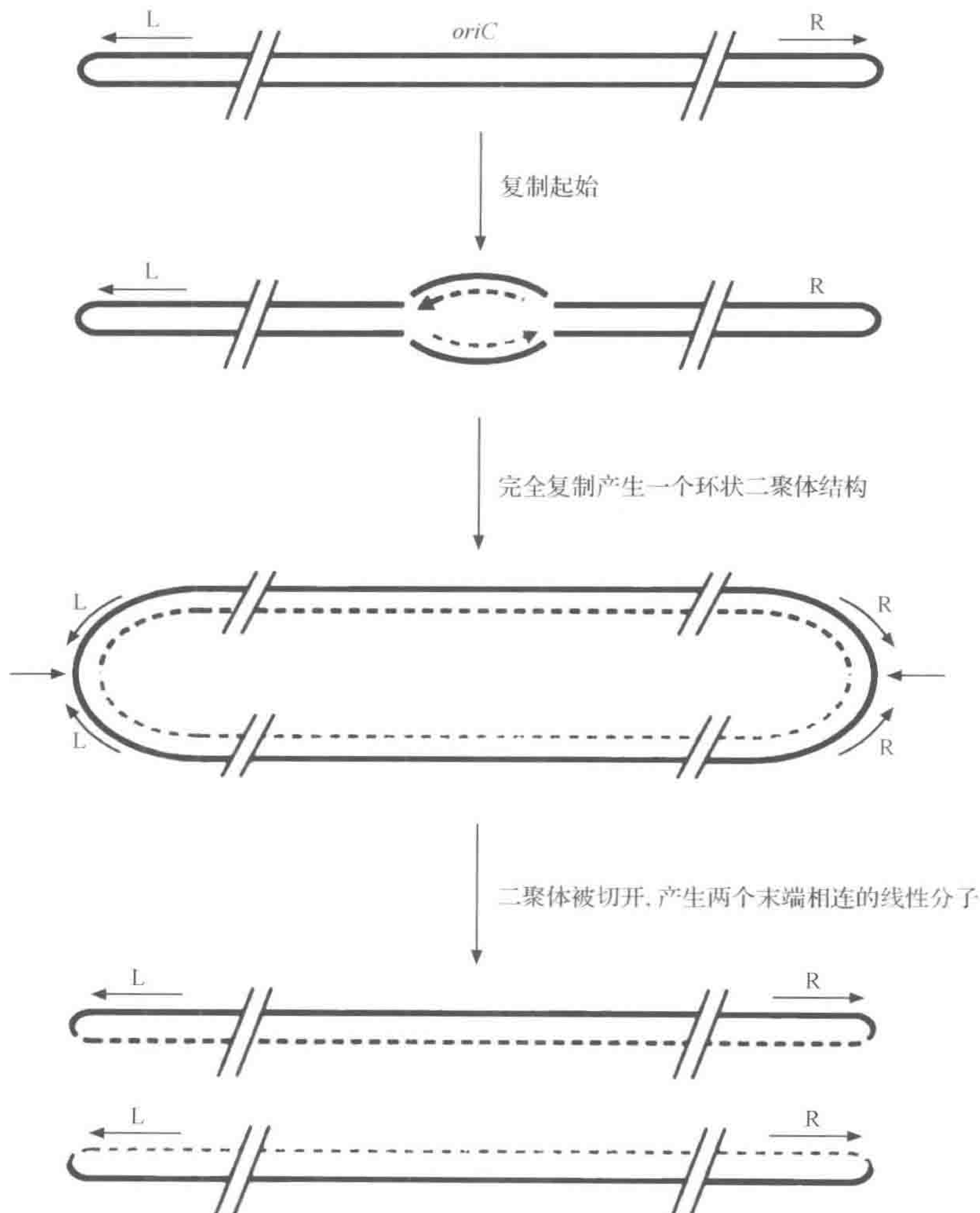


图 5.11　伯疏氏螺旋体线性质粒的复制模式

链霉菌的线性 DNA 两端的复制问题以不同的方式得以解决。一个关键的特点是存在一种共价连接到 DNA 5′端的蛋白质(末端蛋白,TP)。最简单的模型(图 5.12)是这种蛋白质充当 DNA 合成的引物,使线性 DNA 两端复制；而且确实一些线性 DNA 分子(如枯草芽孢杆菌噬菌体 $\varphi29$)的复制以这种方式发生。但是,这并不足以解释链霉菌质粒的复制(甚至解释不了线性染色体的复制)。这些分子在其内部有一个起点,复制以传统方式双向发生。前导链(5′→3′方向产生)被合成至模板末端,后随链将短暂停止,DNA 的一个短的部分未被复制(图 5.13)。在这种情况下,末端蛋白的作用是由于转化的重复序列的存在,通过未复制链二级结构的形成,修补不能复制的片段。

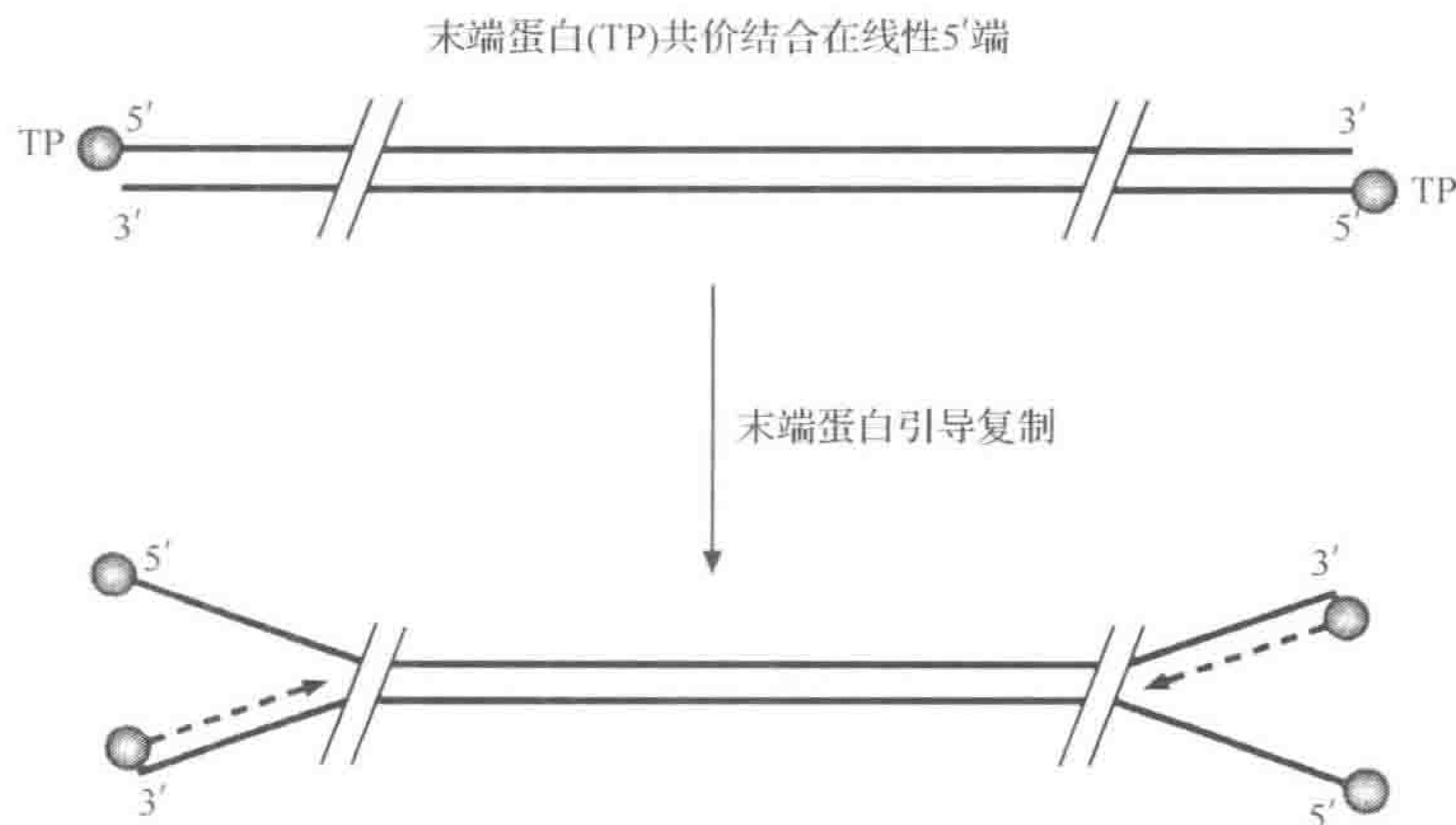


图 5.12　以共价结合蛋白为引物的线性 DNA 的复制

5.2.2　分配

在细胞分裂时,质粒的稳定性要求每一个子细胞都接受质粒的一个拷贝。在这一章的前面部分了解到,通过随机分配,高拷贝的质粒就可以做到这一点。然而,低拷贝的质粒需要更主动的过程确保其发生。

R100 质粒含有一个称为 *par* 基因座(用于分配)的区域,该区域为质粒在细胞分裂时精确分布所必需。这个序列只能以顺式发挥作用,即它必须出现在质粒自身,而不是其他的质粒。非常典型的是(图 5.14),*par* 区包含两个编码蛋白质的基因(*parA* 和 *parB*)和发挥分配机制作用的第三个基因座(*parS*)。与真核细胞的染色体分配机制相类似,*parS* 基因座有时作为着丝点(不同质粒的这些构件名称有所不同,但其特征相似)。*parS* 基因座包含着作为 ParB 结合位点的大量反向重复序列。之后,ParA 结合 ParB,以起始复合物的形式将两个质粒拷贝连在一起。ParA 具有 ATPse 活性,能够形成细丝,其间 ATP 被水解,形成的细丝推动两个质粒拷贝分开,因此当细胞分裂时,每个子细胞都包含了质粒的一个拷贝。

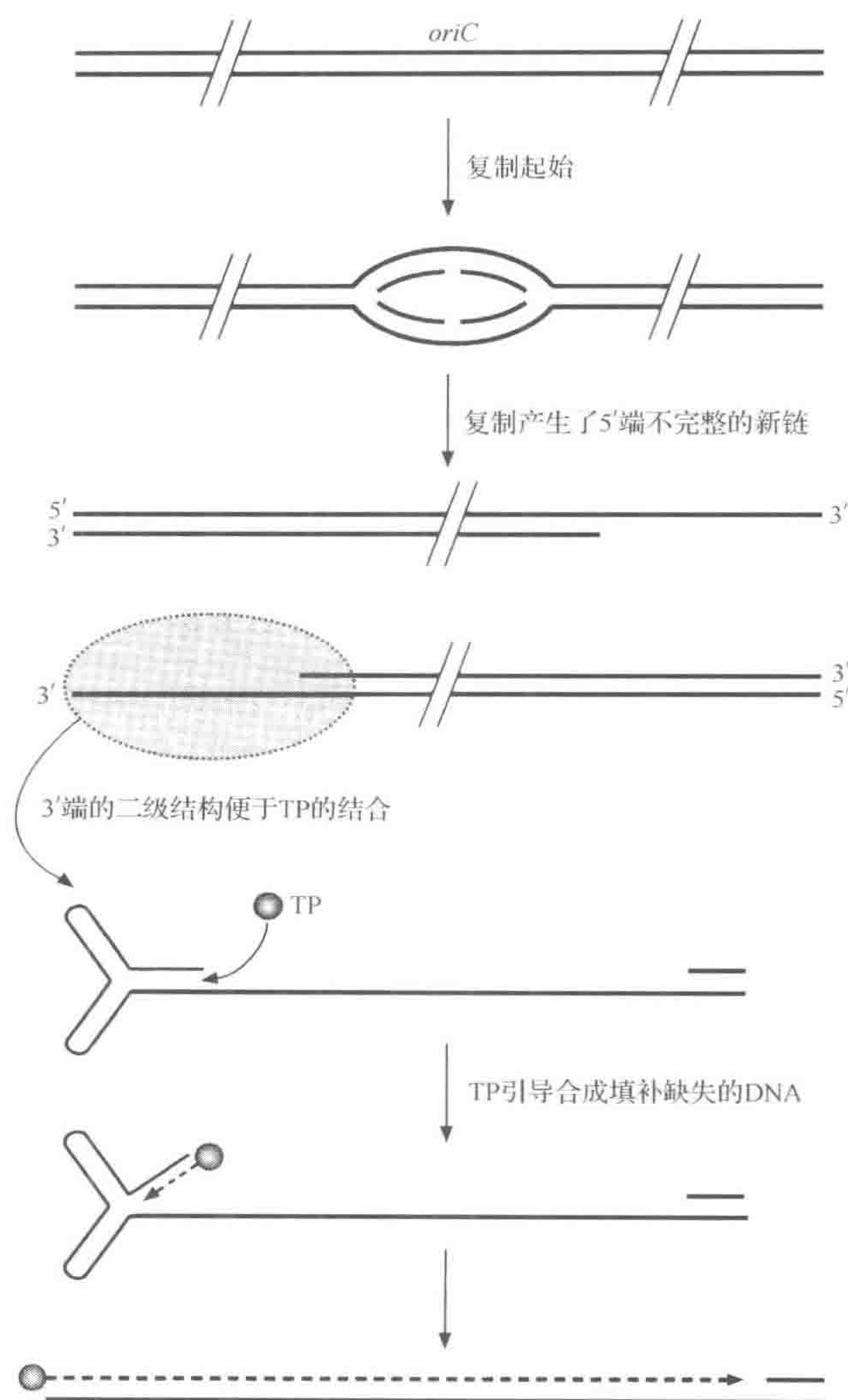


图 5.13 链霉菌线性质粒的复制模式

TP 为末端蛋白(terminal protein)

5.2.3 宿主范围

许多质粒，包括 ColE1 和与它相关的克隆载体，只能在有限的宿主范围内复制。虽然上述机制集中在质粒复制过程中的质粒编码蛋白质的作用，但也有宿主因素的参与，使质粒的复制起点只在关系近的细菌发挥作用。在第 8 章中，将看到一个质粒如何可以在两个不同的细菌中复制，这就是所说的穿梭质粒。

然而，一些质粒具有非常广泛的宿主范围。其中值得注意的是能够在某些革兰氏阳性菌以及大多数革兰氏阴性菌中复制的质粒的 P 群，如 RP4 质粒。RP4 等质粒各式各样，它们不但可以在较大范围的宿主菌内复制，而且也能够通过接合促进其自身在多种细

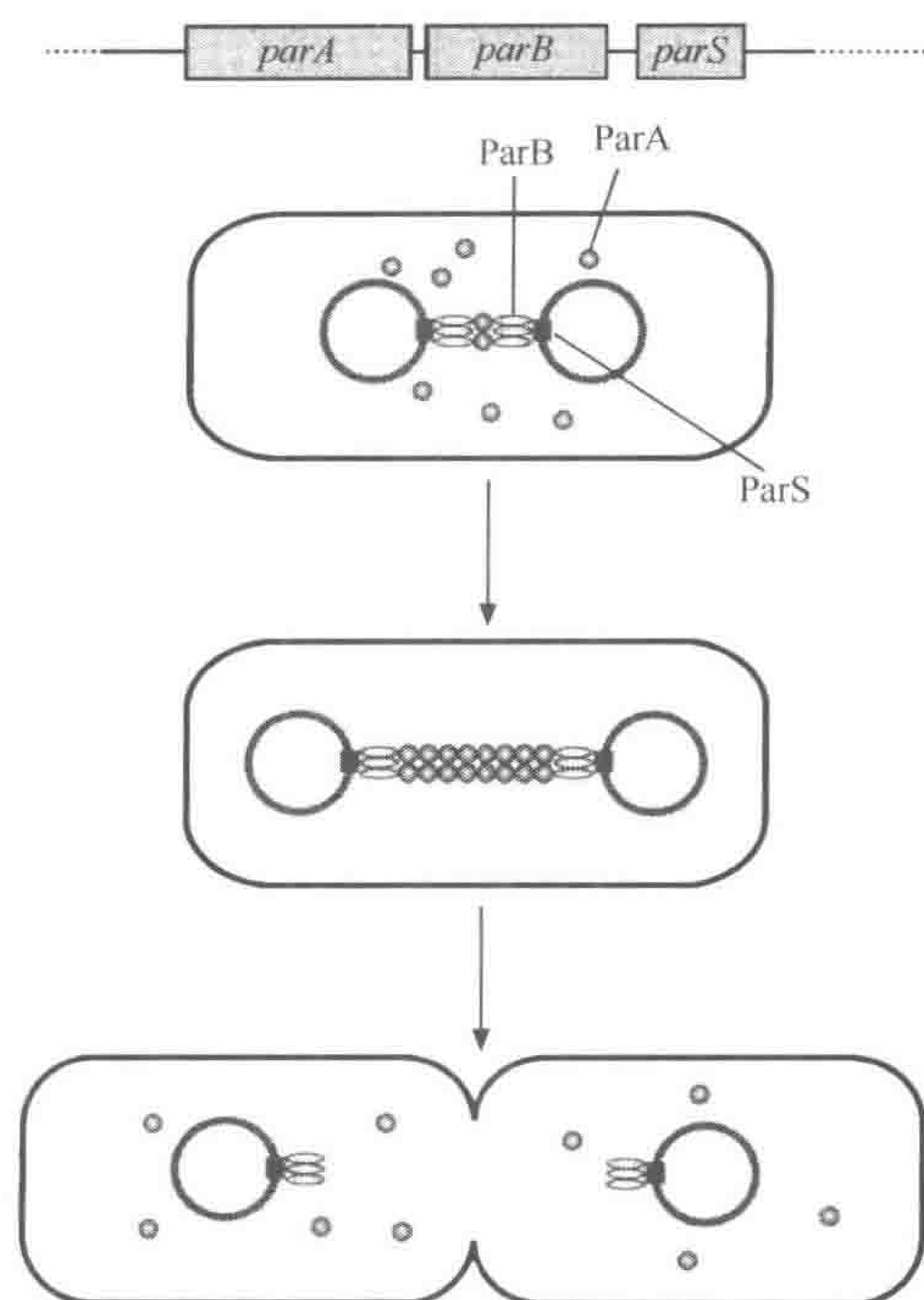


图 5.14　质粒分离系统的模式

ParB 结合到 DNA 的 *parS* 位点上；ParA 结合到 ParB 上，并多聚化形成细丝，推动两个质粒拷贝的分离，从而在细胞分裂时各自遗传给不同细胞

菌之间转移(第 6 章)。宿主范围广泛的质粒不仅在促进广泛多样的细菌之间的基因交流方面发挥了重要的作用，而且也可以成为有用的遗传工具(第 8 章)。

5.2.4　质粒不相容性

细菌可以同时携带几种不同的质粒。然而，并非所有的质粒可以在细胞内稳定共存。有些质粒是不相容的。如果一个细菌细胞开始有两个不相容质粒，一方或另一方丢失的频率非常高。这种现象用来定义质粒不相容的群体，因此，在同一组的两个质粒不能共存。例如，已经遇见过的两个质粒 R1 和 R100，都属于 IncFⅡ群，并且彼此不相容。不相容性的一个原因在于前面所述的复制控制系统，特别是 *copA* 基因的性质。在这两个质粒中，*copA* 基因序列非常相似，产物是可以互换的。因此，R100 的 *copA* RNA 抑制 R1 的复制，反之亦然，从而在一个或其他质粒在细胞分裂时丢失。前面所述的控制质粒复制的 iteron 模式(见图 5.7)，也可能导致不相容性。如果两个不同的质粒携带相关 iteron 序列 (因此 RepA 可以和二者结合)，质粒将连接在一起，复制将被阻止。因此，这两个质粒不能稳定地保持在同一个细胞内。

复制控制机制不是不相容性的唯一原因。有着相似的分配系统的两个质粒也有可能不相容。发生这种情况的一个模型如图 5.15 所示。如果分配装置形成了一个不同质粒的混合对，这可能会导致在细胞分裂时一个质粒的两个拷贝进入相同的子细胞，其结果是

每个细胞都不携带两个质粒。

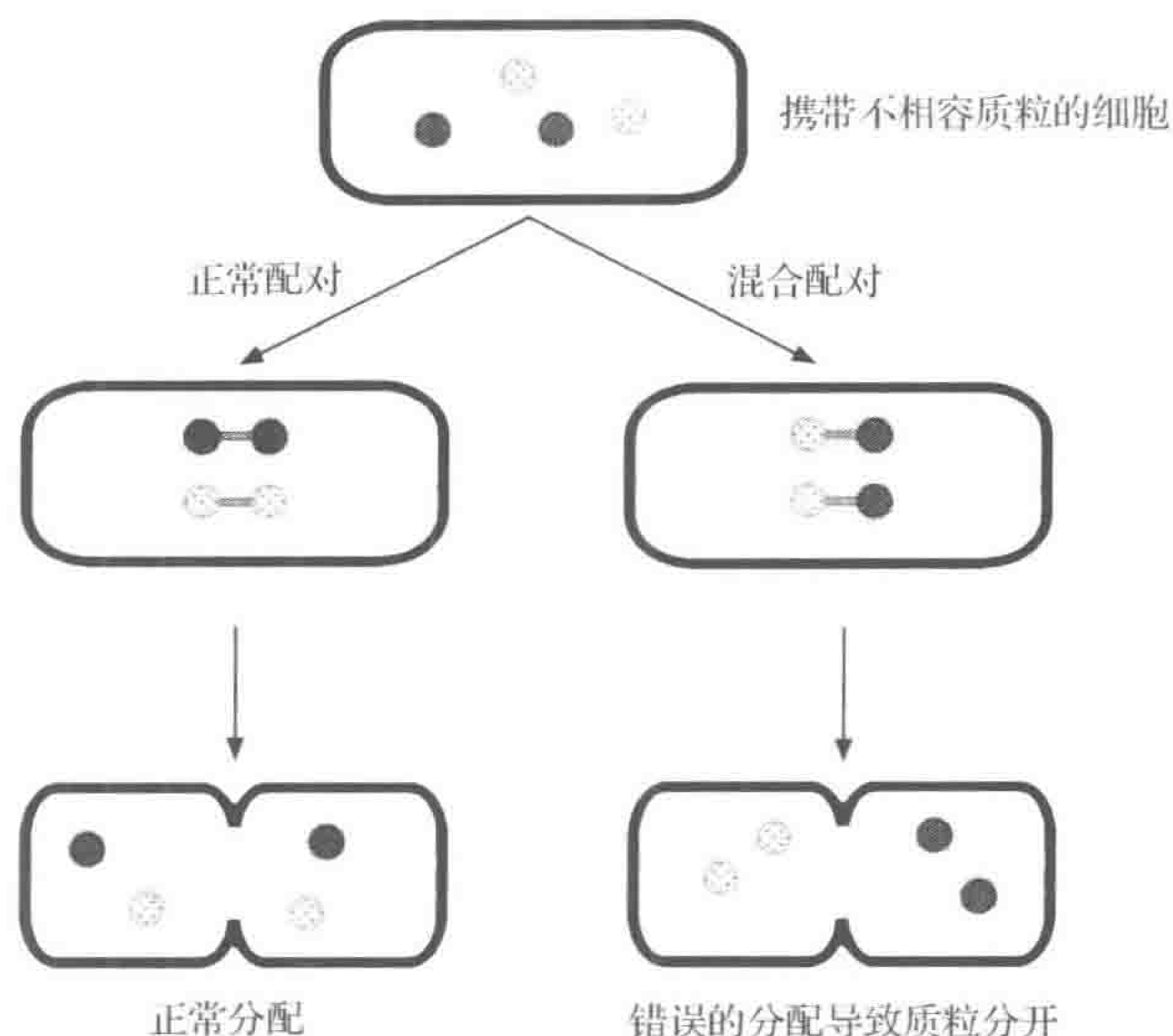


图 5.15 相似的分离系统导致的质粒不兼容性

大量的不相容性群已经得到确定。例如，F 属于 IncFⅠ群，并与如 ColV-K94 质粒不相容。尽管 R100 与 F 在许多方面类似，并彼此相容，正如已经看到的，F 属于 Inc 的不同群，即 IncFⅡ。

低拷贝数质粒的不相容性最显著。对于严格控制的单拷贝质粒，第一次细胞分裂很可能会导致子细胞含有其中一个质粒，而不是二者。对于多拷贝质粒，应用随机分配，不相容性并不明显，与其本身的每个质粒相比，丢失质粒的概率更高。然而，如果两个质粒携带不同的抗性基因，通过抗生素选择的应用，在培养物中保留两个质粒是可能的。

5.3 质粒的稳定性

质粒的特点之一是其不稳定性。与所预计的正常突变过程相比，质粒所具有的特性从一个菌群丢失的频率更高。不同的质粒之间的这种不稳定性的差异程度很大。自然存在的质粒通常(但不总是)相当稳定:选择趋于以该种方式发挥作用，除了分离菌株，以及寻找更倾向选择的更稳定的质粒，发现不稳定的质粒更难。

另外，人工构建的质粒往往是明显不稳定的。这通常只是一个实验室范围的麻烦，但当工业用菌株携带这类质粒时，会成为一个代价非常昂贵的问题。

作者将在下面讨论 3 种与质粒稳定性的概念相关的现象——质粒完整性、在细胞分裂时分配、生长率差异。

5.3.1 质粒完整性

完整性是指维持质粒的结构。在这方面，即使是自然存在的质粒也可能是“不稳定”的。由于存在重组热点会呈现一种丢失基因的倾向。特别是重复序列的存在，如转座子

或插入序列(第7章),可能会由于重复序列之间的重组导致缺失或倒位。在图5.16中,质粒B缺失了含有卡那霉素和氯霉素抗性基因的区域。氨苄青霉素抗性的测试可能会使大家相信质粒相当稳定。然而,通过对其他抗生素的抗性测试,或通过确定质粒的大小(更为常用),可以确定缺失。由于质粒大小保持不变,并可能保留了原来所有的表型特征,对发生了重排(如质粒C的倒位)的质粒进行鉴定是比较困难的,但可以通过质粒测序进行证实。在第7章中,将进一步讨论质粒或染色体上的倒位可能产生的影响。

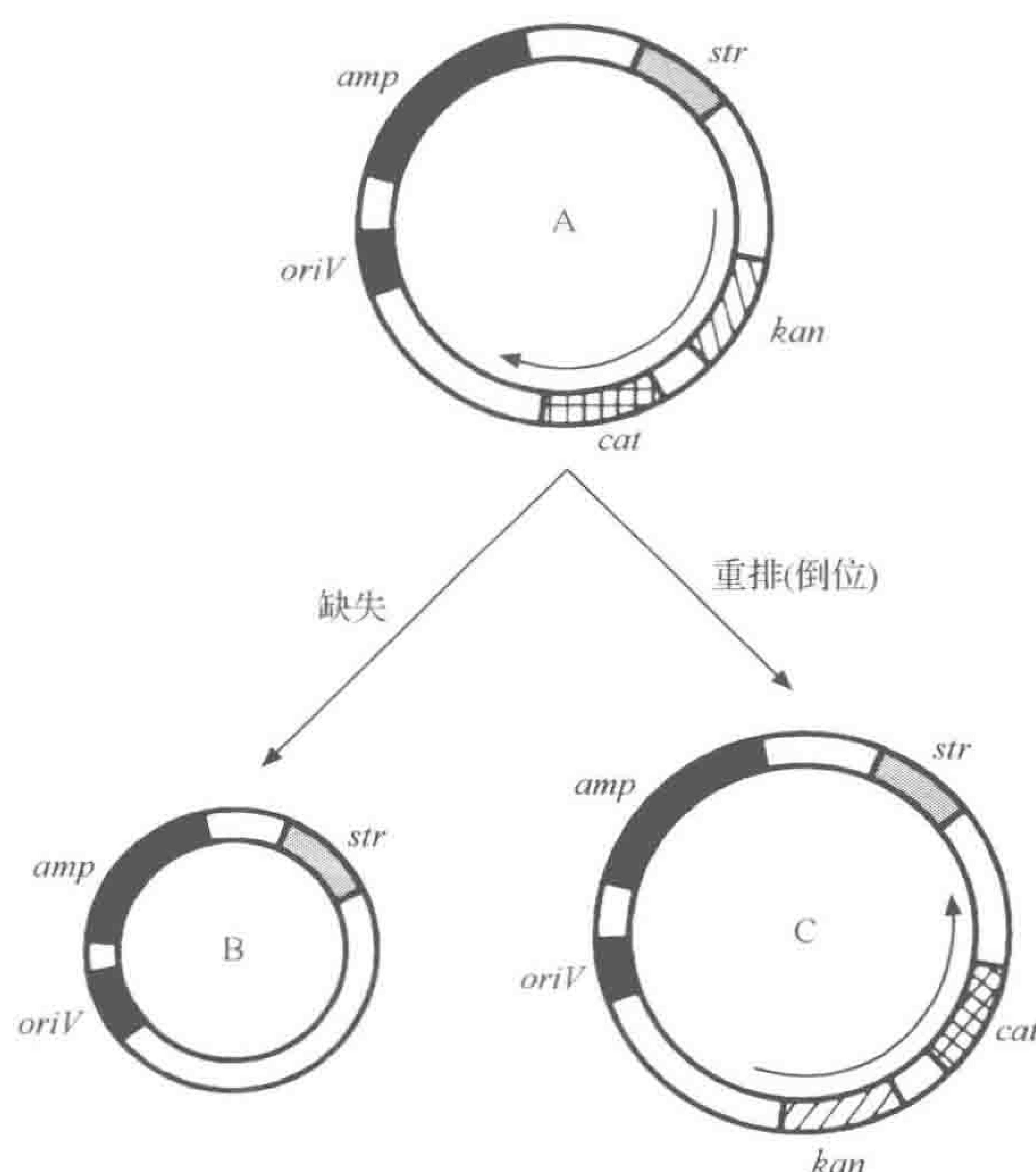


图5.16　质粒的完整性

列举两条质粒结构发生改变的途径。质粒B缺失了*cat*和*kan*基因,而质粒C的该区域发生了倒位。*amp*、*str*、*kan*和*cat*分别抗氨苄青霉素、链霉素、卡那霉素和氯霉素。*oriV*为复制起点

5.3.2　分配

如前所述,如果质粒要在培养物中保持存在,在细胞分裂时的正确分配是必需的(图5.17)。虽然高拷贝质粒主要依靠两个子细胞之间的随机分布,这可能会受到在复制过程中质粒形成多聚体结构倾向,以及单体间重组的影响。此外,由于多聚体包含多个复制起点,其复制效率比单体更高。这可能会导致所谓的"二聚体灾难",其中二聚体的比例和更高的多聚体,在一定程度上增加了对质粒持续维持的威胁。

在这里考虑的是质粒是否以单体、二聚体或多聚体存在为什么非常重要。控制质粒拷贝数的模型的工作基本上是通过计算复制起点数,而不是分子数目。因此,一个具有30个拷贝数的质粒,将以30个单体,或15个二聚体,或10个三聚体的形式存在。因此,当涉及细胞分裂时,一个相当大比例的多聚体的存在将减少质粒的有效拷贝数,并因此增加其中一个子细胞不能接收质粒的一个拷贝的可能性。

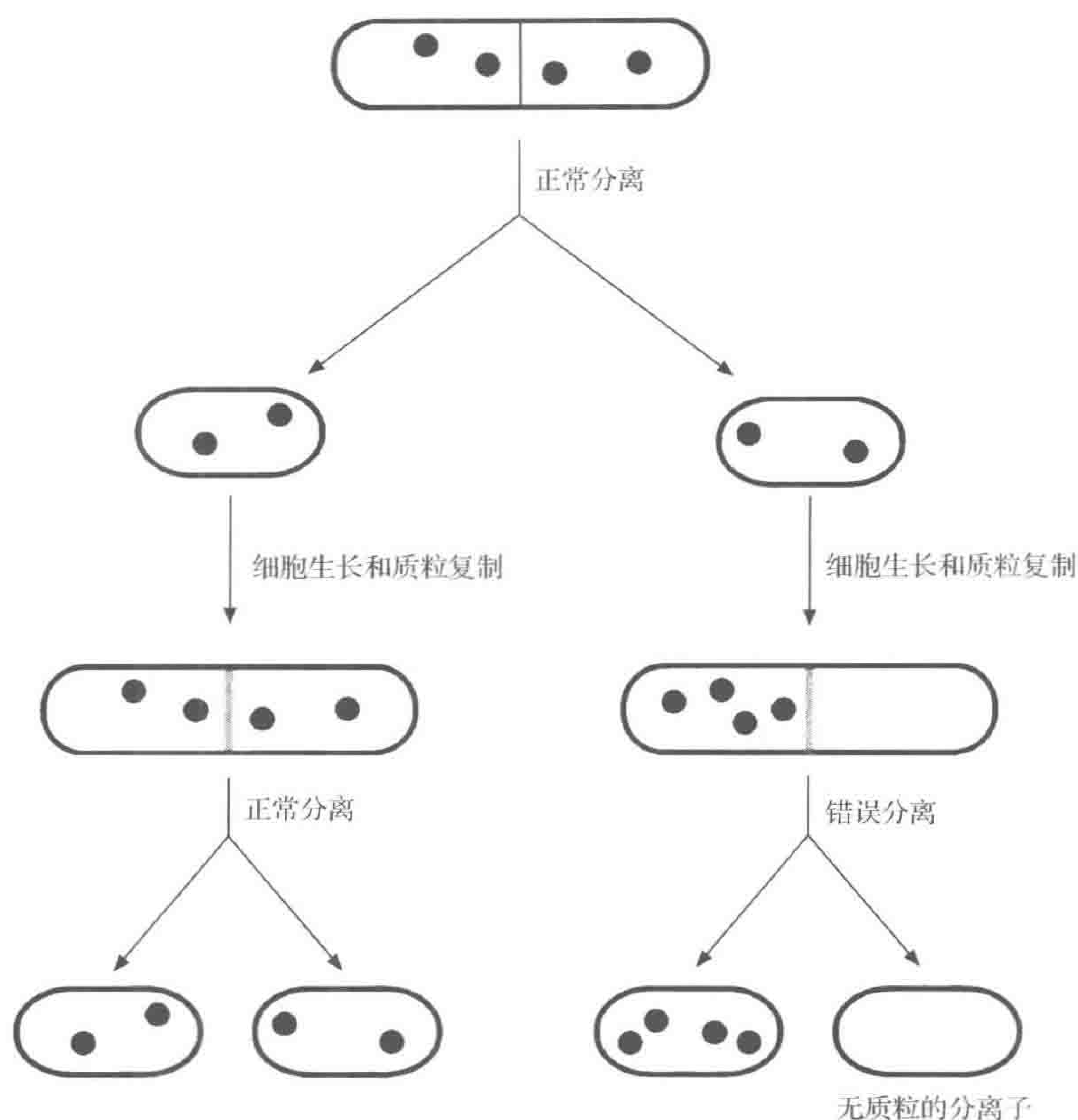


图 5.17 通过失败的分离使质粒分离

对抗这种效果的主要机制包括位点特异性重组。例如，ColE1 包含一个被称为 *cer* 的位点，该位点是 XerC 和 XerD 蛋白的作用靶点。在二聚体中，有 *cer* 序列的两个拷贝，Xer 蛋白将促进它们之间的重组、断裂两个 DNA 分子，彼此交叉，然后再连接，从而将二聚体结构分离为两个单体(图 5.18)。源于 ColE1 *cer* 基因缺失的明显的不稳定性使该系

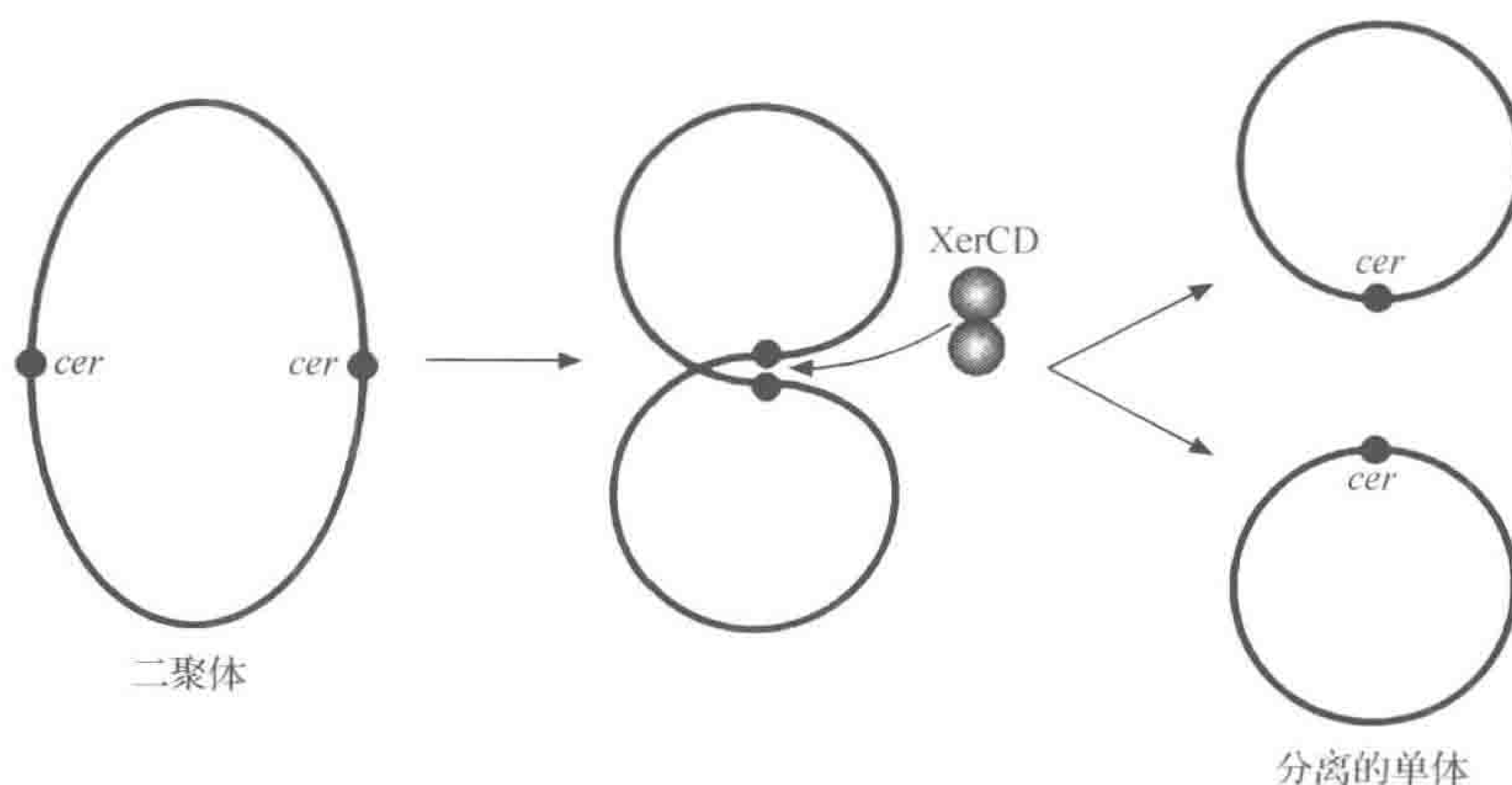


图 5.18 通过位点特异性重组分离质粒二聚体

在二聚体质粒的 *cer* 位点，由 XerCD 引起位点特异性重组，使其分离成质粒单体

统的重要性得以显示。

质粒依赖性

作为一个有活性的分配机制，一些质粒以能够杀死任何丢失质粒的细胞的能力补充其分配系统（分离后杀死）。这些系统由两部分组成：一个稳定、长寿命的毒素和一个充当其解毒剂的不稳定因素。只要质粒存在，就会产生解毒剂，但如果质粒不存在，不再产生解毒剂，并且迅速降解。这种毒素仍然存在，但是，在没有解毒剂存在的情况下，细胞会被杀死。

一个例子是包含了一个含有 *ccdA* 和 *ccdB* 两个基因的操纵子的 F 质粒（图 5.19）。因为它会干扰 DNA 复制（通过其对 DNA 促旋酶介导的超螺旋的作用），CcdB 蛋白质是有毒的，而 CcdA 蛋白（更不稳定）拮抗 CcdB 的作用。

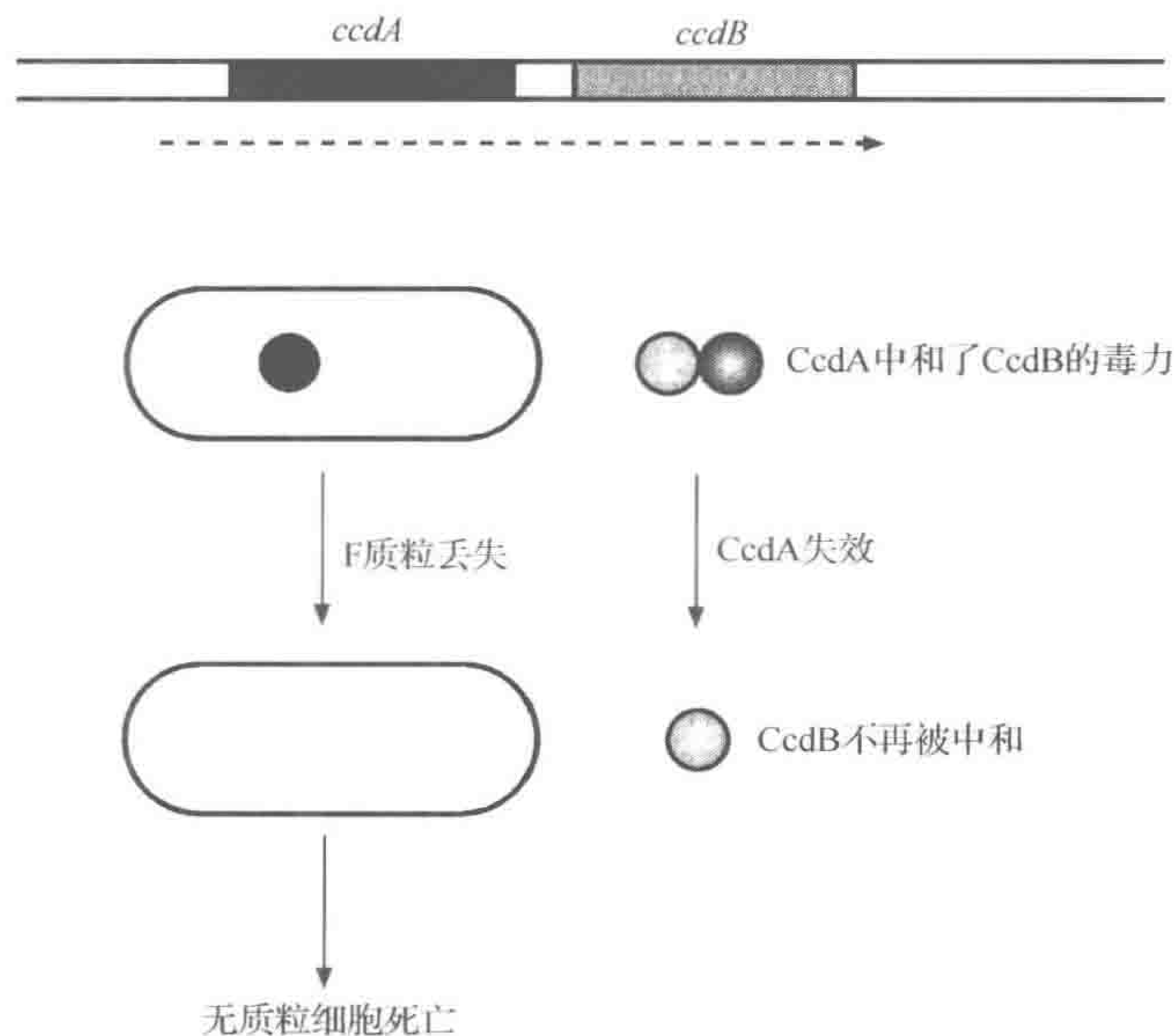


图 5.19　*ccdAB* 基因在持有 F 质粒中的作用

F 质粒编码的 CcdB 蛋白是有毒的，但可被 CcdA 中和。如果质粒丢失，CcdA 被迅速摧毁，而使细胞暴露于稳定的 CcdB 蛋白的毒力效应中

在无质粒分离子中，CcdA 产物被破坏，而稳定的 CcdB 蛋白仍然存在，并杀死细胞。在多种无关的质粒上，发现了相似的系统。

在某些情况下，解毒剂通过调节毒素的表达发挥作用。例如，质粒 R1 携带 *hok* 基因（杀死宿主的意思），该基因编码一个因为其对于细胞膜的作用而有毒的小的多肽。在含有质粒的细胞，与 *hok* mRNA 前导区互补的一个反义 RNA 分子（*sok* 为杀死抑制的意思）抑制了 *hok* mRNA 的翻译：从 DNA 的同一部位的相反方向转录。*sok* RNA 分子迅速衰变，但 *hok* mRNA 非常稳定。因此在无质粒分离子中，*hok* mRNA 仍然存在，在 *sok* RNA 不存在的情况下，翻译产生毒素。这个系统的简化模型如图 5.20 所示。

在细菌染色体上也发现了类似的系统。其中一些是整合元件（如噬菌体、转座子或整

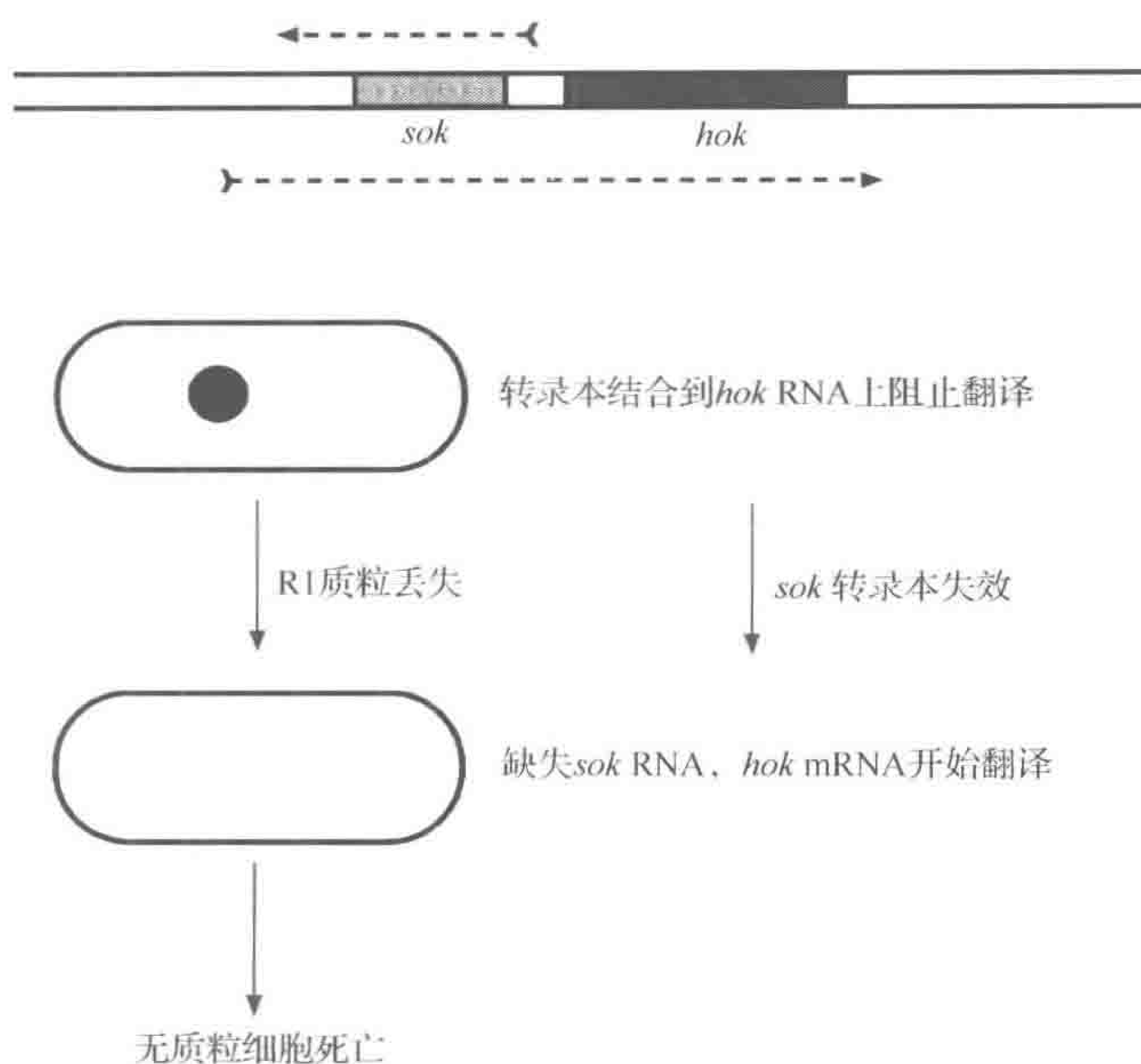


图 5.20 *hok* 和 *sok* 基因在持有 R1 质粒中的作用

R1 *hok* 基因产物是有毒的,但其翻译被反义 *sok* RNA 所阻止。如果质粒丢失,*sok* RNA 迅速失效,*hok* mRNA(更稳定)被翻译,导致细胞死亡

合质粒) 的一部分,并具有确保整合元件稳定性的目的。其他一些是正常的染色体基因。大肠杆菌基因组包含 7 个毒素/抗毒素对,其中研究最多的是 MazEF 系统。MazF 是一种毒素,是一种剪切 mRNA 的核糖核酸酶,导致细胞死亡,而 MazE 是一种与毒素结合并使之灭活的抗毒素。虽然该系统的功能还不完全清楚,但其与质粒依赖性系统是不同的。一个证据来自如果翻译被阻止,则两种蛋白质都不能产生的事实,但因为毒素比抗毒素更稳定,一些细胞会因为毒素的作用而死亡。这样的后果是,在一定的应激条件下,细胞会发生程序性细胞死亡。对于单个死亡的细胞,这显然是不利的,但对于一个整体来说,这可能是有利的。例如,在饥饿条件下,一部分细胞的死亡可能有利于促进剩余细胞的生存。这意味着,应该将一种细菌培养物视为一个类似于多细胞的生物体,而不是作为一个完全独立的个体的集合。这也可见于在细胞之间发生的信号转导。在第 9 章中,会遇到的细菌——黏液球菌(*Myxococcus xanthus*)就存在一个明确的多细胞阶段。

总之,细胞分裂时质粒的稳定性存在于①活性分配 (低拷贝数质粒)或随机分配(高拷贝数质粒);②二聚体或多聚体分离为单体质粒;③分离后杀死。

5.3.3 生长率差异

第三个参数也很重要,即携带质粒的细胞和那些不携带质粒的细胞之间存在生长率差异吗? 如果没有差异,那么在细胞分裂时的分配失败只会导致无质粒细胞的比例缓慢增加。另外,即使分配失败发生的很少,生长率的显著差异会导致质粒的迅速消除。

因为质粒的复制及其基因表达所产生的代谢负载,生长率差异会增大。大多数野生

型质粒的这些要求都小,因此对生长率的影响也小。另外,遗传改造的质粒经常保持非常高的拷贝数并且以非常高的水平表达某一特定产物。因此对细胞的负担非常大,不稳定的问题会比较尖锐。

5.4 与表型相关的质粒

表型变异往往是质粒所涉及的第一个表现,也就是说,某些菌株具有一个丢失频率高于预期突变的特定特征,而且该丢失是不可逆的。通过各种处理,如吖啶橙或较高温度下生长,即"固化"或质粒消除,可以增加丢失质粒的比例,必须强调的是,"消除"和"质粒丢失"并不意味着从一个特定的细胞中物理移动质粒;通过干扰复制和(或)质粒分配以增加无质粒分离子的出现比例,这个过程也得以进行。

这方面的证据应结合质粒 DNA 琼脂糖凝胶电泳检测(图 5.21,并参见框 2.3)。染色体 DNA 会呈现一个比较弥散的带形,因为提取过程中的随机片段化,而任何质粒 DNA 都主要基于其大小在确定的位置形成单独的、清晰的条带。分子越小,它将移动越快。在染色体 DNA 前面发现小质粒,而较大的质粒实际移动的要比染色体 DNA 慢。这似乎有点出人意料(染色体要比最大的质粒还要大得多),但要记得染色体会断裂成线性片段,而质粒会以完整的环状分子存在。因此,它们之间没有直接的可比性。

图 5.21　从细胞中抽提质粒的琼脂糖凝胶电泳示意图

质粒的构象也将影响其移动性。通常,完整的质粒是一个共价闭合的超螺旋环,所以在相同分子质量下,其与线性分子的移动速度不同,并快于缺口开放环状分子(在尝试确定质粒大小时要注意这一点)。除了质粒的这 3 种主要形式(超螺旋、线性和缺口),可能存在移动速度更慢的二聚体和更高的多聚体。由于一个质粒可以出现几条带,确定一个给定菌株中有多少质粒就会非常困难。

为了更深入地研究质粒,对其进行纯化是必要的。从其他细胞成分中分离 DNA 是比较容易做到的。然而,从染色体 DNA 中分离一个质粒并不是很简单的。最常用的方法是依赖超螺旋环状质粒和染色体的线性 DNA 片段的不同特性。然而,对于较大的质

粒，或质粒的其他形式，如本章前面提到的单链和线性质粒 DNA，这些方法则不一定适合。

质粒 DNA 表型的不稳定性和其物理特性的结合为质粒的参与提供了证据。通过证明菌落特性的丢失同样也发生在质粒上，这种假设得以加强。但证据仍然是推论性的。如果需要毋庸置疑的证据，最好的办法是将质粒引入一个没有质粒的细胞（通过接合或转化，第 6 章），看其是否会导致表型的相应变化。应用可选择的标记，如抗生素耐药性，这样的实验是很容易做到的。如果质粒携带一种耐药基因，通过将细菌铺板到含有该抗生素的培养基，就可以检测到获得质粒的细胞。通过这种方式，可以检测到极其罕见的事件，例如，10^8 个细胞中检测到一个抗性转化接合子（或转化子）。其他的特点就不会如此容易被检测到了。如果想知道质粒是否携带了为毒力所必需的基因，这可能是相当困难的。

6 基因转移

早在孟德尔之前，人们就已经熟悉了动植物通过有性繁殖使其性状得到重组这一理论。不仅人类的个体特征代表了其父母（或祖父母）特征的重组，而且在动植物的育种中，已经应用这一现象组合了不同品系（植株）优良性状的新品种（系）长达几个世纪。在不表现有性生殖的生物体（如细菌）中又该如何利用这一现象呢？

细菌不仅在实验室存在遗传信息的交换，在自然界亦然，其3种交换机制截然不同。

1. 转化：细菌摄入其周围环境中游离的DNA分子。
2. 接合：涉及DNA分子从一个细胞转移到另一个细胞。
3. 转导：是由侵染细菌的病毒（噬菌体）介导的基因转移。

并非所有的菌种都会表现出3种基因转移方式。接合在革兰氏阴性菌中最常见，而在革兰氏阳性菌中也会发生，如链霉菌属（*Streptomyces*）和链球菌属（*Streptococcus*）。尽管有些重要的菌种可以自发地发生转化，但很多其他菌种则必须先对细胞进行一些人工的预处理才能发生转化，因此这些生物体可能不会自发地发生转化。

这些机制与真正的有性繁殖比较，主要有两个方面的不同：这些机制与繁殖没有联系；亲本对个体的遗传贡献不相同。亲本分别被称为供体细胞和受体细胞，重组后代在很多特性方面与受体更相似。

本章主要关注细菌间的自发的基因转移（水平基因转移），以及自发的基因转移对细菌进化的重要性。实验室所用到的基因转移机制，尤其是基因修饰、基因扩增和遗传分析等方面的应用将在第8～10章讲解。

6.1 转化

在某种意义上，转化的发现，即细菌细胞对DNA的摄取的发现开启了对今天所熟知的细菌遗传学和分子生物学的研究。1928年，Fred Griffith在研究肺炎链球菌（*Streptococcus pneumoniae*）时发现，将非致病菌与灭活的致病菌的提取物共同培养可使非致病菌获得致病性。16年后，Avery、MacLead和McCarty证明了该转化物质为DNA，从而建立了DNA是细菌的遗传物质的理论。转化在某些菌种的遗传分析中很重要，由于其在基因克隆方面所扮演的重要角色，近年来更加受到关注。

对于肺炎链球菌来说，细菌自发性地变得能够摄取外源DNA。这种自发的转化现象在枯草芽孢杆菌（*Bacillus subtilis*）、流感嗜血杆菌（*Haemophilus influenza*）（与肺炎链球菌一样）中研究的较多，而且在一段时间内，认为只限于这几个菌种及相关的菌种。现在人们已经知道其广泛存在。在淋球菌（*Neisseria gonorrhoeae*）的抗原变异中，转化作

用尤其广泛。该种抗原的变异是由于 *pil* 基因的转入而实现的。*pil* 基因是编码一种菌毛主要蛋白质亚基的基因，而细菌通过菌毛吸附于上皮细胞。尽管已证明能够自发转化的菌种的数量还相当有限，但在很多其他细菌中也会发生(尽管水平较低)。在本文中，记住这点很重要，那就是环境中存在大量的细菌，它们的繁殖速度极快。因此，即便是一个很小的事件，其作用在进化历程中也会显得尤为重要。所以，转化在很多菌种的进化中可能扮演了重要的角色。

不同菌种转化的详细过程各不相同，但存在一些普遍性的规律。感受态通常出现在细胞生长的特定时期，一般是对数生长晚期，也就是细胞进入稳定期时，这可能是其对细胞浓度而非细胞生长时期的反应。例如，枯草芽孢杆菌(*B. subtilis*)中一些参与感受态形成的基因也参与了早期阶段的孢子形成。此阶段感受态的形成不仅与营养消耗有关，而且还与特异的分泌物(感受态因子)的累积有关，该分泌物通过一个双组分的调控系统刺激其他感受态所需基因的表达。因为感受态因子的水平依赖于细胞浓度，所以，只有高细胞浓度时才会出现感受态。这是一个群体感应形式，如在第 3 章中所描述的，周围介质中的细菌浓度控制了单个细胞的反应。

随着感受态的形成，多条双链 DNA 片段结合到细胞表面的受体上，但其中只有一条能够进入细胞。在某些菌种中，通过获得一个短的菌种特异性片段，该过程选择性地只摄入来源于同一菌种的 DNA。例如，脑膜炎球菌(*Neisseria meningitidis*)的 DNA 摄取就依赖于一个 10bp 的特异性摄取序列的存在。在脑膜炎球菌基因组中含有 2000 个该片段的拷贝，该现象在其他菌种的基因组很少发生。与此类似，在流感嗜血杆菌(*H. influenza*)的基因组中的一个 29bp 的片段有助于该菌的转化，此片段在其基因组中大概有 1500 个拷贝。因此，只有来自相同种的片段才能有效转化这些菌种。

另外，枯草芽孢杆菌(*B. subtilis*)和肺炎链球菌(*Str. pneumonia*)几乎可以摄取任何线性 DNA 分子；但 DNA 的摄入还仅仅是个开始。要使细胞被稳定转化，摄入的 DNA 必须复制并遗传。在此讨论的是染色体 DNA(而非质粒)，摄入的片段只有与宿主染色体 DNA 发生重组后才能复制。重组需要转化片段与受体染色体间具有同源性。对于来自其他菌种 DNA 的转化，这点并不是绝对的障碍。只要与染色体的某些区域有足够的相似性，这些 DNA 片段就可以与受体染色体发生重组。在分类关系上越相近，其转化就越有可能具有相似性。在这方面，非常具有实际意义的一个例子是肺炎链球菌对青霉素的耐药性，这是由于编码青霉素靶酶基因的一部分被抗口腔链球菌的相应片段所取代。

由于线性 DNA 分子效果最好，而不是在基因修饰中常用到的环形质粒 DNA，自然的转化对基因的修饰作用是有限的。为了将外源的 DNA 导入宿主细胞，人们使用了各种技术诱导细菌细胞进入感受态。另外，可以通过将细胞与 DNA 的简单混合物放于高压电场中，从而使 DNA 进入细胞(该过程被称为电融合)。尽管电融合所用的转入 DNA 的机制不同，但也具有细胞摄入裸露 DNA 的特性，故也被称为转化。这种方法将在第 8 章中详细讲解。

6.2 接合

接合是指 DNA 从一个细菌直接转移到另一个细菌。尽管在生物体也会发生染色体

DNA 的转移，但在多数情况下，接合涉及质粒 DNA 的转移。与其他的转基因方式相比，接合是一种从一个亲本(供体)到另一个亲本(受体)单向的 DNA 转移。

在实验室条件下，最简单的接合情况是，将两个不同菌株混合，共培养一段时间使其发生接合，然后放入选择性培养基，使两个亲本型的菌株都不能生长，而带有双亲本基因的接合子却能生长。如图 6.1 所示，一个菌株(供体)是带有抗氨苄青霉素的质粒，而另一个菌株不含质粒，但其染色体 DNA 是具有抗萘啶酸的突变型。经混合共培养，取样涂板于含有两种抗生素的培养基上进行培养。两个亲本型的菌株在该培养基上均不能生长，故所观察到的菌株是供体质粒转移到受体细胞的菌株。尽管这是个概率相当低的事件，但通过这两种抗生素提供的强有力的选择，只要在 10^6 个受体中有 1 个发生了接合，就可以很容易地筛选出转移质粒的菌株。

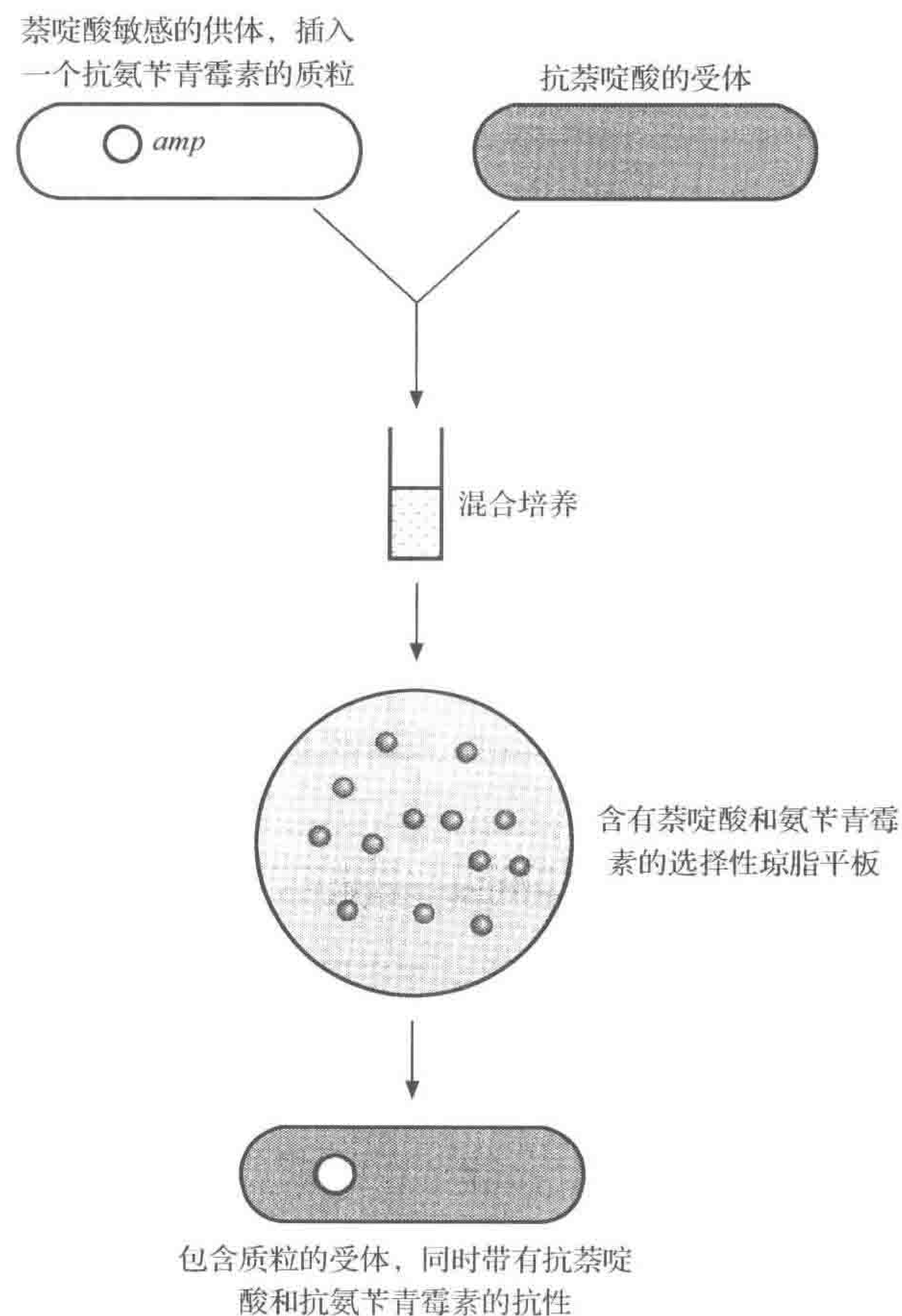


图 6.1 抗性质粒的接合型转移

供体菌株对萘啶酸敏感，且带有抗氨苄青霉素的质粒。受体由于染色体的突变，对萘啶酸具有抗性，但对氨苄青霉素敏感。将其混合培养后，接种于含有萘啶酸和氨苄青霉素的选择性平板上，筛选出接受了质粒的受体(转接合子)，简要起见，细菌的染色体未显示出

在肠杆菌及一些其他的革兰氏阳性菌(如弧菌和假单胞菌等)中，接合较为常见。革兰氏阳性菌的几个属具有很明显的接合系统特性：包括链霉菌(是生产抗生素的主要菌，

具有重要的商业价值)，乳酸链球菌(因其在乳制品工业中各方面的应用具有重要的商业价值)，还有一些具有重要医药价值的细菌，如粪肠球菌(*Enterococcus faecalis*)(将在本章后面进行讨论)。

接合最重要的意义在于可使质粒从一个菌株转入另一个菌株。因为接合并不限于同种菌株的成员之间，故为遗传信息跨菌种间分类界限的转移提供了一种更广阔的途径。一种结果是处于肠道的质粒能够被传送给病原体，使之能够对各种抗生素产生抗性。

6.2.1 接合的机制

交配个体的形成

在多数情况下，接合的发生依赖于供体菌株含有的质粒携带的促进 DNA 转移的基因。在大肠杆菌及一些革兰氏阴性菌中，供体细胞表面携带着称为菌毛的附属物。这些菌毛在结构上有很大的差别。例如，F 质粒特有的菌毛长、细且具有弹性。而 RP4 菌毛则短、粗且具有刚性。菌毛与受体细胞表面的受体蛋白接触，使供体细胞与受体细胞形成交配个体(图 6.2)。然后继续拉进两个细胞间的距离使其直接接触，并形成一个通道或孔隙。DNA 经此通道从供体进入受体。有趣的是，这种通道在蛋白质的分泌系统中很常见(Ⅳ型分泌，见第 1 章)。有些细菌利用该分泌系统将蛋白质毒素直接转入宿主细胞。

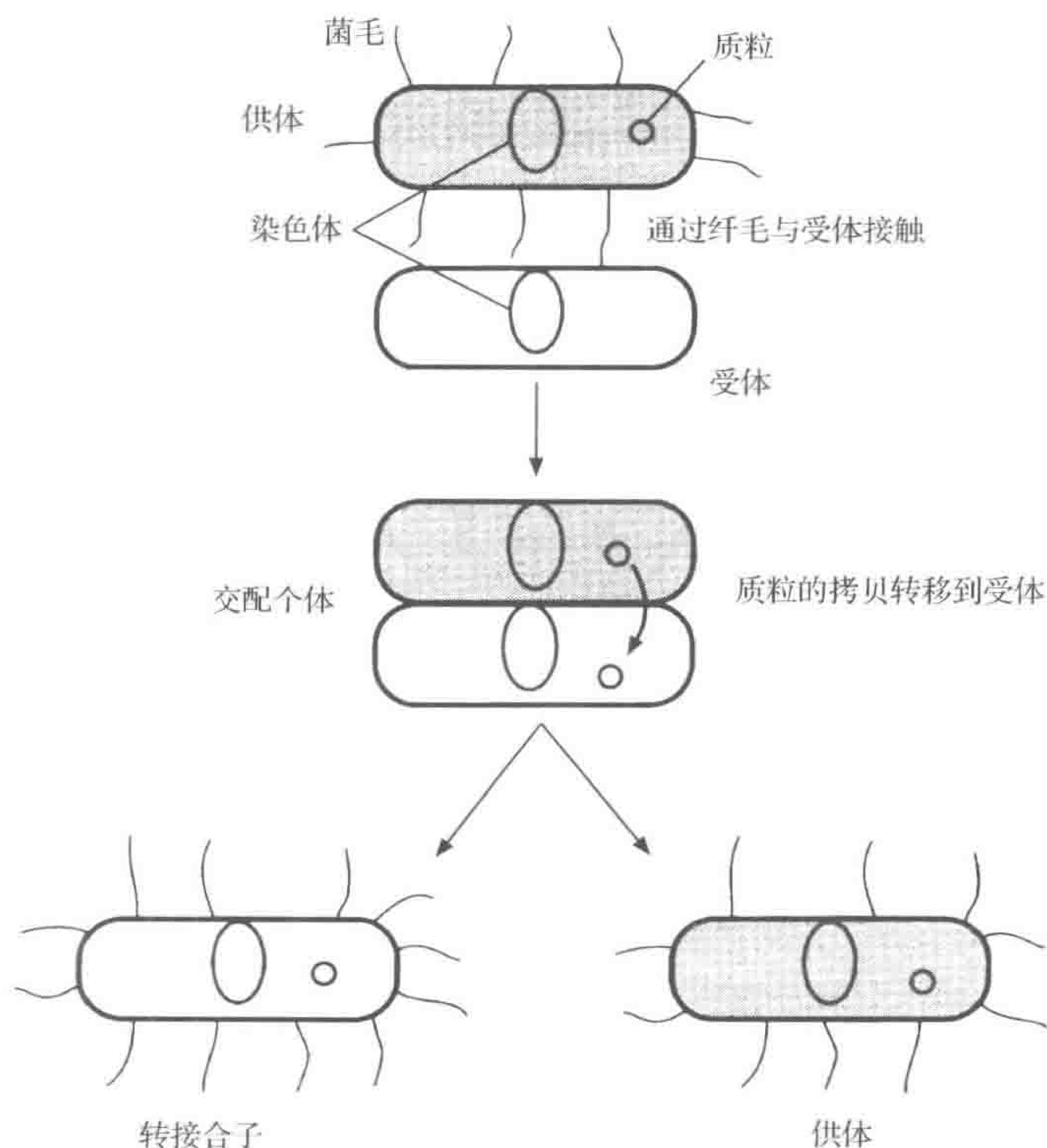


图 6.2 DNA 通过接合转移

在本章的后面部分，将对革兰氏阳性菌较重要的接合机制进行讨论。

DNA 的转移

质粒 DNA 从供体到受体的转移(图 6.3)起始于一个蛋白质,该蛋白质可以在称为转移起始点的 *oriT* 位点的特异性位点切断单链 DNA。一个质粒编码的解旋酶将质粒 DNA 解螺旋,然后带切口的单链 DNA 从 5′端开始向受体转移,直到切口时结束。同时,切口链游离的 3′端以滚环式复制的形式不断地延长并取代被转移的 DNA,其滚环复制类似于单链质粒 DNA 及噬菌体 DNA 的复制(见第 4 章,第 5 章)。切割蛋白依然吸附于转移 DNA 的 5′端。受体 DNA 的合成使被移入的单链 DNA 变成双链 DNA 分子。

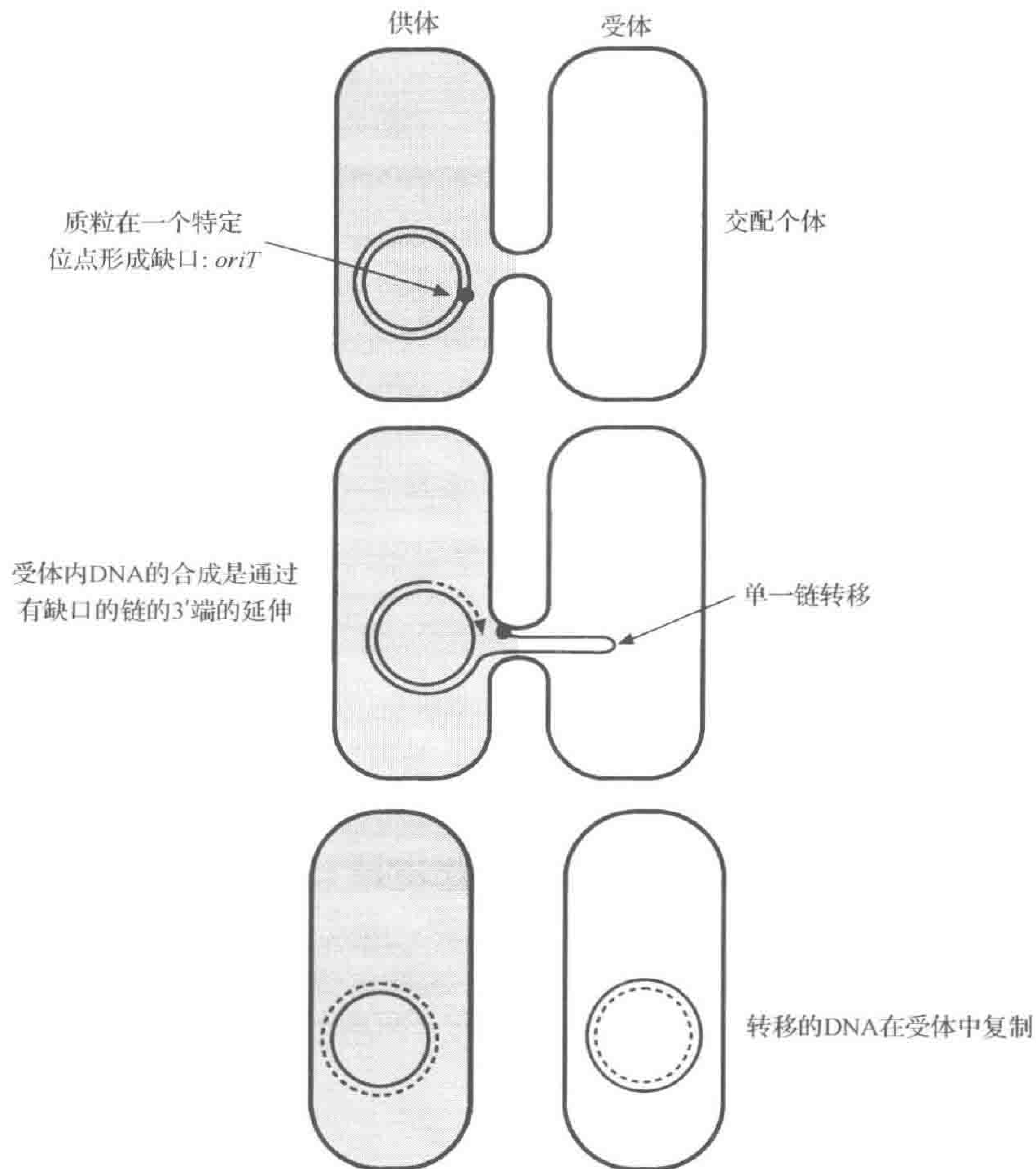


图 6.3　质粒 DNA 通过接合转移的机制

为简要,图中只画出了质粒

值得关注的是,这是一个复制过程。所以,尽管所说的质粒从一个细胞转移到了另一个细胞,但在真正意义上,转移的是质粒的一个拷贝。供体菌株仍然拥有该质粒的一个拷贝,而且可以转给其他受体。由于受体细胞也有一个质粒的拷贝,也可以将质粒的一个拷贝转移给其他细胞,这点也值得关注,其结果是,通过群体的混合,质粒呈流行性扩散。

迁移作用和染色体的转移

并非所有的质粒都能够自主地转移到另一个细胞，有这种能力的质粒称为接合型质粒。在有些情况下，接合型质粒可促使第二个与其共同存在于供体细胞内的非接合型质粒的转移。这种情况不是随机发生的，也并不是所有非接合型的质粒都可以被迁移。

为了理解迁移作用，以 ColE1 质粒为例。ColE1 是一个非接合型的质粒，但它具有 *mob* 基因(见图 5.3)，该基因编码一种特异性的核酶，用于上述迁移模式中 DNA 链的切割。*oriT* 位点(又称 *bom*)也是 Mob 酶切割 DNA，使其形成切口的位点。因为 ColE1 不携带配对所需的基因，所以是非接合型的。由于供体中另一个接合型质粒的存在使供体与受体能够配对，然后 ColE1 质粒可利用自身的机制完成 DNA 的转移。

有些可迁移的质粒不带有 *mob* 基因。迁移作用依赖于接合型质粒的 Mob 酶对被迁移质粒的 *bom* 位点的识别能力。只有当两个质粒在分类上非常相近时，这种作用才奏效。另外，*bom* 位点对于迁移来说是必需的，在基因修饰中这点非常重要。当从一个质粒载体中将 *bom* 位点剔除后，可以确保被修饰过的质粒不被转移到其他菌株(见第 8 章)。

在多数情况下，从供体转移到受体的 DNA 都仅由质粒 DNA 的一个拷贝组成。然而，有些类型的质粒可以使染色体 DNA 发生转移。其中第一个被发现、并被人们所熟知的是大肠杆菌的 F 质粒(F 是致育因子)，但类似的系统在其他菌种也存在，在绿脓杆菌属尤为明显。在下述的一些情况中，这种转移涉及接合型质粒整合于供体的染色体内(所以染色体作为质粒的一部分也被转移)。然而，在很多情况下，染色体 DNA 转移的发生并不与质粒 DNA 有稳定的联系，也可能是通过一种与非接合型质粒的转移类似的机制来完成。

当质粒 DNA 通过接合作用从一个细胞转移到另一个细胞时，整个质粒都被转移。与此相比，染色体 DNA 的转移并不涉及一个完整染色体拷贝的转移。转移所需要的时间是其中一个原因。转移效率比正常的复制要低，转移整条染色体需要 100min(在大肠杆菌)，交配个体很少能保持这么久。40kb 的质粒 DNA 才相当于染色体长度的 1%，所以，人们推测质粒 DNA 的转移可以在 1min 内完成。染色体 DNA 的不完整的转移意味着其在受体细胞内不能形成一个完整的复制子。这些片段要复制、遗传必须与宿主染色体发生重组。在重组过程中，通常会替代受体 DNA 的相应部分。

6.2.2 F 质粒

F 质粒最初是在证明遗传交换的实验中被发现的。该实验是将大肠杆菌的两个营养缺陷型菌株混合后培养于基本培养基，结果只有重组型细菌能够生长，从而证明了遗传重组。很早就有实验证明：重组体都源自亲本菌株之一，涉及从供体(雄性)到受体(雌性)信息转移。供体菌株带有 F 质粒(F^+)，而受体菌株是 F^-。当时令人们感到奇怪的是，这个系统的一个特点是将 F^+ 和 F^- 质粒共培养后可使雌性变成雄性的结果，这当然是由于 F 质粒转移的结果。与 F^+ 供体发生的染色体片段的转移相比，F 质粒的转移发生的频率很高。

Hfr 菌株

由于一种供体菌株的发现，接合作用在遗传分析上的应用被大大加强。该菌株中的染色体 DNA 的转移更为普遍。这些 Hfr(高频重组)菌株源于质粒整合进入细菌染色体。Hfr 菌株另一个特点是染色体 DNA 的转移，从一个限定的点开始向一个特定的方向进行。F 质粒的插入位点决定了转移的起始，插入质粒的方向决定了转移的方向。当染色体 DNA 中含有一个整合的 F 质粒时，可以通过如重复标记环状大分子来阐明这一点(见图 6.3)。转移起始于整合的质粒的 *oriT* 位点，而这导致细菌染色体的一个拷贝也被转移，并非仅仅是质粒。F^+供体的基因转移或多或少具有随机性，因为其转移不起始于染色体上限定的位点。一部分染色体 DNA 和基因的有序转移相结合，使接合成为绘制细菌染色体图谱的重要手段(见第 10 章)。

F 质粒的整合与切离以及 F′质粒的形成

通过质粒上的一段序列与染色体上的某个位点发生重组，发生了 F 质粒的整合。整合后，F 质粒在整合位点呈线性存在。因为整合质粒的末端位点间的重组可能导致质粒以独立的环状分子从染色体上切离，所以整合是可逆的。然而，不精确的切离是可能的，即重组发生于不同位点。如果发生这样的情况，将会导致质粒 DNA 带有小部分的细菌 DNA，这样就形成了 F′质粒。例如，质粒整合于与乳糖操纵子相邻的位点，其不精确的切离可导致 F′*lac* 质粒的形成。在 F′*lac* 质粒，乳糖操纵子整合在质粒中，当质粒转移时，乳糖操纵子也随之转入受体菌株，使受体能够发酵乳糖。这是一种质粒可从细菌获得额外的基因，并将其转入另一菌株或菌种的机制。

F′质粒可以用于创建部分二倍体，即除了染色体拷贝，在质粒上具有一个特定基因的一个拷贝。在第 3 章中，讨论了应用部分二倍体研究乳糖操纵子的调控，尤其是用于区分反式作用的调节基因(这些基因影响不同分子的基因表达)和只影响其所吸附的基因。这些实验仍然是相关的，但现在可以使用通过体外生产的重组质粒。

6.2.3 其他细菌的接合

上面对接合的描述主要适用于大肠杆菌、假单胞杆菌(*Pseuclomonas*)等革兰氏阴性菌。许多革兰氏阳性菌，从杆菌属到球菌属也都含有可通过接合转移的质粒，而且，在很多例子中，其 DNA 的转移机制与上述描述的非常相似。然而在其他方面，它们却存在着很大的差别。一般来说，接合性转移所需基因的数目在某些菌种可以少到只有 5 个，比革兰氏阴性菌所需的至少 20 个要少得多。因此，革兰氏阳性菌中的接合型质粒非常小。所需基因少的一个原因可能是它们不需要编码菌毛的基因，这可能至少部分地反映了革兰氏阳性菌在细胞壁结构方面的特性，即与革兰氏阴性菌相比缺少外膜结构。

接合机制研究的较清楚的革兰氏阳性菌是肠球菌，主要是粪肠球菌(*E. faecalis*)，图 6.4描述的是其简单模式，如带有 pAD1 这样的质粒的供体菌株能够自我生长，则转移所需的质粒基因会被一种抑制蛋白 TraA 的作用所关闭。这些基因的激活需要可以分泌扩散肽的受体细胞的存在(在本例中，扩散肽为 cAD1)。扩散肽具有类似激素的作用，这

些肽是通过对多肽前体的蛋白水解切割而产生的。供体细胞带有一个质粒编码的蛋白TraC。cAD1被转运出细胞壁后与TraC结合。cAD1一旦进入供体细胞，将会与TraA阻遏物结合并激活它，使转移基因表达。转移基因的一个产物——聚集物Asa包裹供体细胞后，将导致供体细胞与受体细胞聚集、配对成为一体。质粒的转移则随之发生，其转移机制与前述相似。

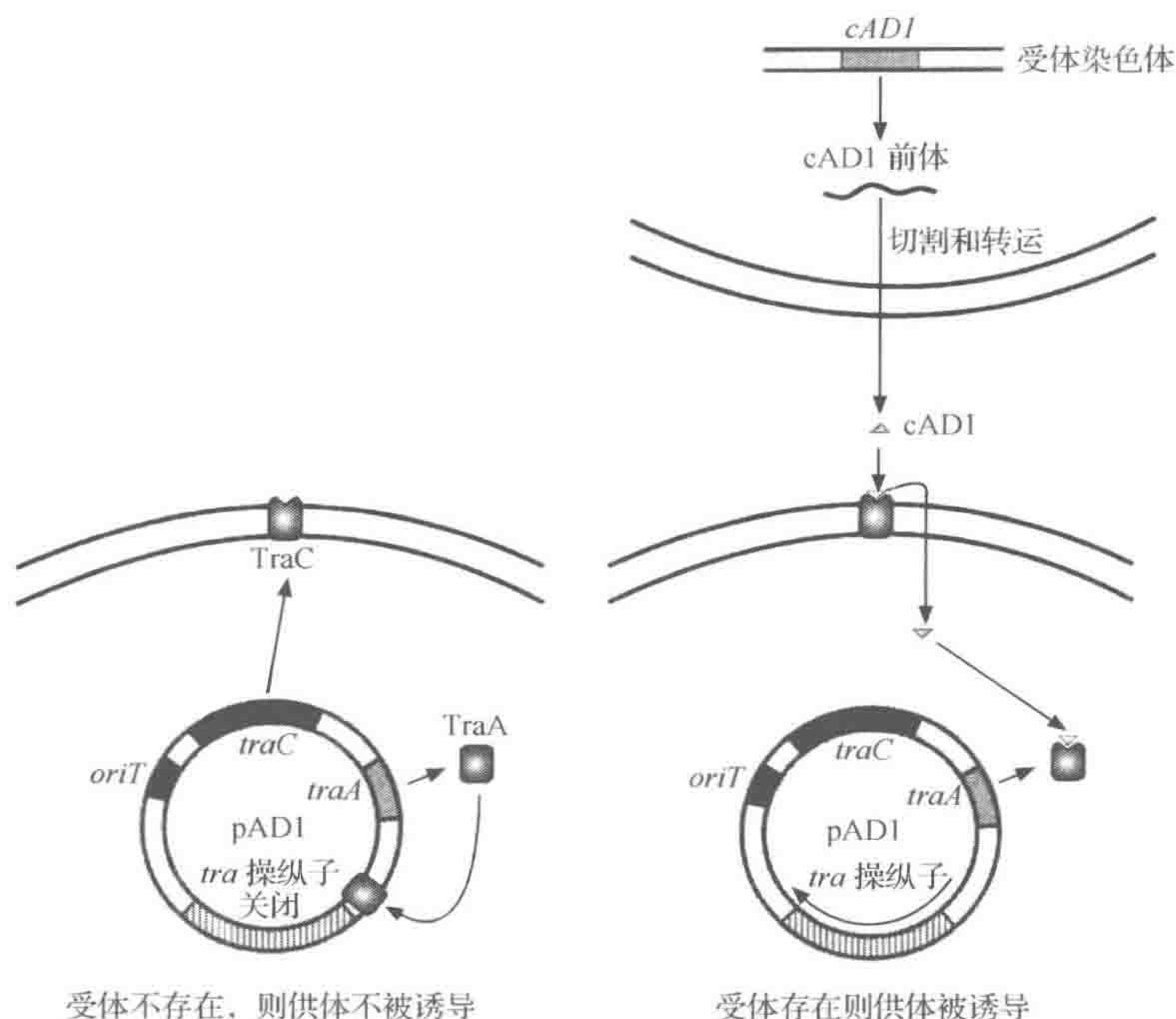


图 6.4 肠球菌中接合型质粒调控的简单模型

在缺少受体的情况下，转移基因被TraA蛋白关闭。受体产生一种肽即cAD1，可以使TraA蛋白失活，从而使接合得以进行

转移后，转接合子(转入质粒的受体细胞)会发生更多的事件，来抑制这种激素的进一步的作用，所以，这个细胞将不再作为pAD1的受体。然而，这个细胞还可以继续分泌其他激素，这些激素对编码不同受体蛋白的不同质粒是特异的。

这个系统的一个优点是含有质粒的细胞不表达转移基因，除非其邻近有受体细胞。这不仅降低了细胞的新陈代谢负荷，而且还意味着，它们不会表达可能会被宿主免疫系统识别的表面抗原(如菌毛)。

接合型转座子

一般来说，接合是以质粒介导的，粪肠球菌(*E. faecalis*)是个特例。有些粪肠球菌菌株含有名为Tn*916*的转座子。转座子将在第7章中详细讲解。但现在暂时需要知道的是转座子是一种可以从一个基因位点转移到其他基因位点的可移动的遗传元件。Tn*916*与其他转座子的区别在于，它具有通过接合从一个细胞转入另一个细胞的能力。

接合型转座子，如Tn*916*，与质粒不同的是，它们被复制并作为染色体的一部分被遗传。与质粒不同，它们的复制没有稳定的、独立的复制方式。然而，对其转移方法的进一

步研究证明，接合型转座子与质粒的转移有许多相似之处(图 6.5)。尤其是 Tn*916* 含有一个转移起始点(*oriT*)，这与许多质粒中发现的 *oriT* 很相似。转移的第一步是在转座子编码的酶(Int 和 Xis)的作用下从染色体切离转座子。转座子编码的酶(Int 和 Xis)与 λ 噬菌体整合与切离的转移有关(见第 4 章)。Tn*916* 切离后形成一个环形分子，该分子除了一点之外与质粒都很相似，这一点就是没有复制起点，因此不能够以正常方式复制。然而，由于其含有一个转移起始位点 *oriT*，并携带有接合型转移所需要的 *tra* 基因，可以被转移到一个受体细胞。

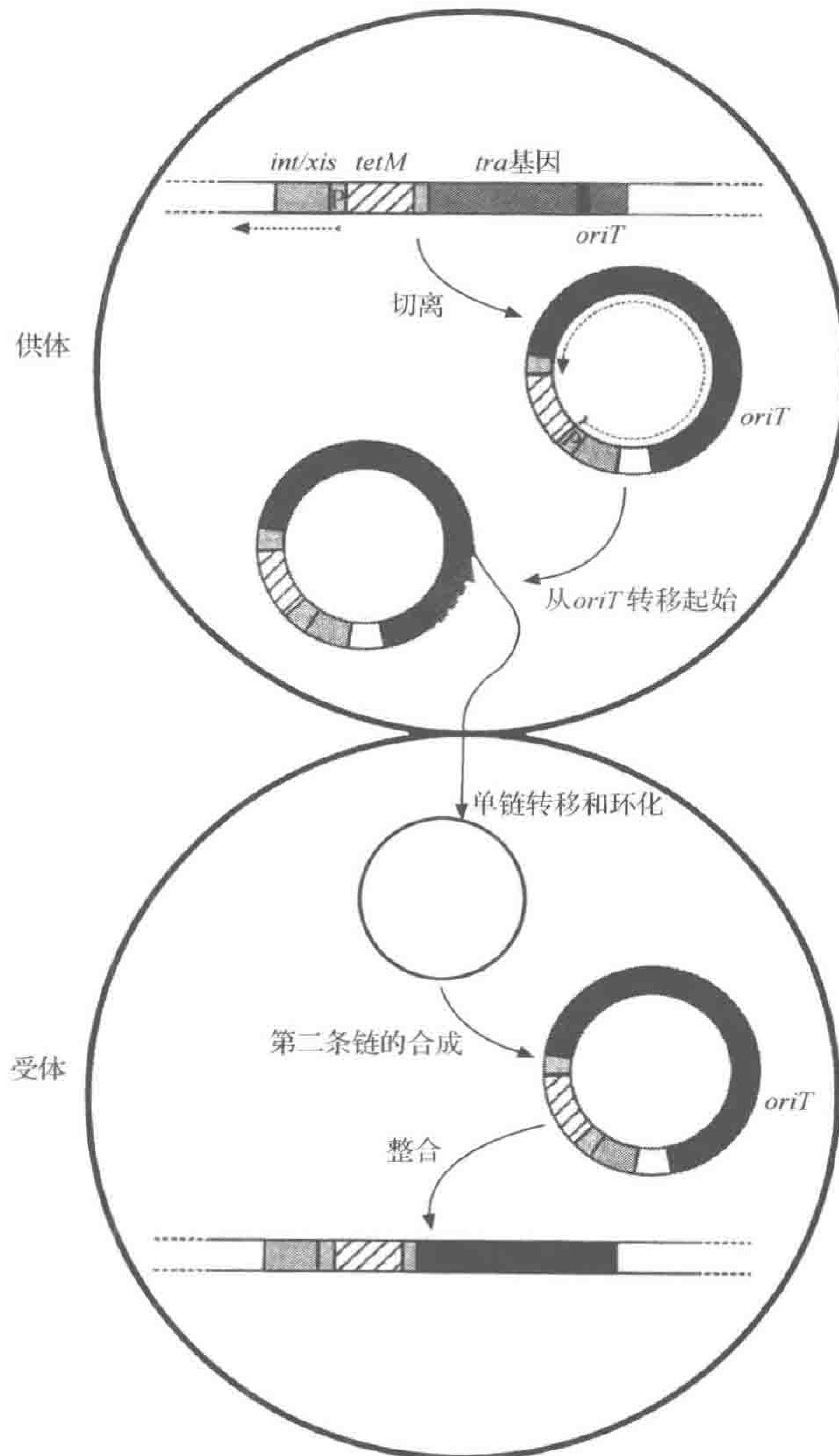


图 6.5 接合型转座子 Tn*916* 的转移

转座子在酶(Int 和 Xis)的作用下切离染色体，并环化。这使得 *tra* 基因得以表达，接合型的转座从 *oriT* 位点开始。在受体细胞中，转入的 DNA 分子形成环形，复制产生第二条链，形成转座子整合进入受体染色体

与在前面所描述的质粒的转移系统相似，Tn*916* 的转移包括单链 DNA 的合成，从转移起始位点 *oriT* 开始，将被取代的链转入受体细胞。被转移的单链环化，转变成环形双链等，偶尔还可在整合酶的作用下插入受体染色体。

Tn*916* 另一个值得关注的特点是，如果在没有从染色体上切离时发生转移，可能会认为染色体也被转移，还伴有不完全的转座子转移。转移如果自 *oriT* 位点开始，则会在到转座子的其余部分之前，围绕着染色体作用。这似乎不会发生，原因是 *tra* 基因表达所需的启动子朝向转座子的左手末端(图 6.5)，背向 *tra* 基因。在线性的整合状态中，*tra* 基因不会被表达。然而，当从染色体上切离转座子并环化后，可使启动子位于 *tra* 基因转录的正确方位。环形中间体会表达这些基因，而在整合形式却不能。这确保了转移系统只有在切离状态下才被激活。

Tn*916* 是一个接合型转座子家族的原型。尽管在革兰氏阴性菌[如拟杆菌(*Bacteroides*)]也会遇到相关的元件，这些转座子在革兰氏阳性的球菌分布尤其广泛。在许多转座元件中，包括 Tn*916*，由于转座子可以转入其他菌种或属，接合型转移都是混杂的。因此可以确信，接合型转移在将遗传材料(包括抗体-抗性基因)传播到整个细菌王国中起到了重要的作用。尤其值得注意的是，其中的许多转座子，包括 Tn*916*，携带有在许多细菌广泛存在的四环素耐药基因(*tetM*)，说明它们在该基因的传播中发挥了重要作用。

6.3 转导

转导是噬菌体介导的遗传材料的转移，其关键步骤是在噬菌体裂解生长过程中将 DNA 包装进入噬菌体头部(见第 4 章)。对噬菌体 DNA 来说，这个过程一般是高度特异的。但在有些噬菌体中也会偶尔出错，使细菌 DNA 片段(产生于噬菌体对宿主染色体的降解)被错误地包装，导致噬菌体颗粒中带有细菌基因组的部分片段(图 6.6)。这些转导颗粒能够感染受体细胞，因为吸附和注入 DNA 所需的信息被噬菌体颗粒的蛋白质所携带，与其所含的核酸无关。转导的 DNA 片段则被注入新的宿主细胞内。

并非所有的噬菌体都能够发生转导。有效转导型噬菌体的基本要求是：感染应导致染色体 DNA 的适度降解，形成合适大小的片段，并且要在正确的包装时间，而包装过程的特异性相对较低。

在某些情况下，被转导的 DNA 是一个细菌质粒，注入的 DNA 分子能够被复制并遗传。在更多情况下，被装入转导颗粒的是染色体 DNA 的片段，在受体细胞中，它们将不能被复制。如果它们想要复制并遗传，与其他基因转移机制的例子相似，则必须插入到受体染色体内(通过同源重组)。

这个过程被称为一般性转导(与特异性转导相反，见下面内容)，实质上任何一个基因都有同等的被转导的机会。

6.3.1 特异性转导

如在第 4 章所述，有些噬菌体(如温度敏感噬菌体)能够形成一种称为溶源的生长状态，在这种状态下，噬菌体基因的表达及噬菌体的复制被抑制。在很多情况下，前噬菌体

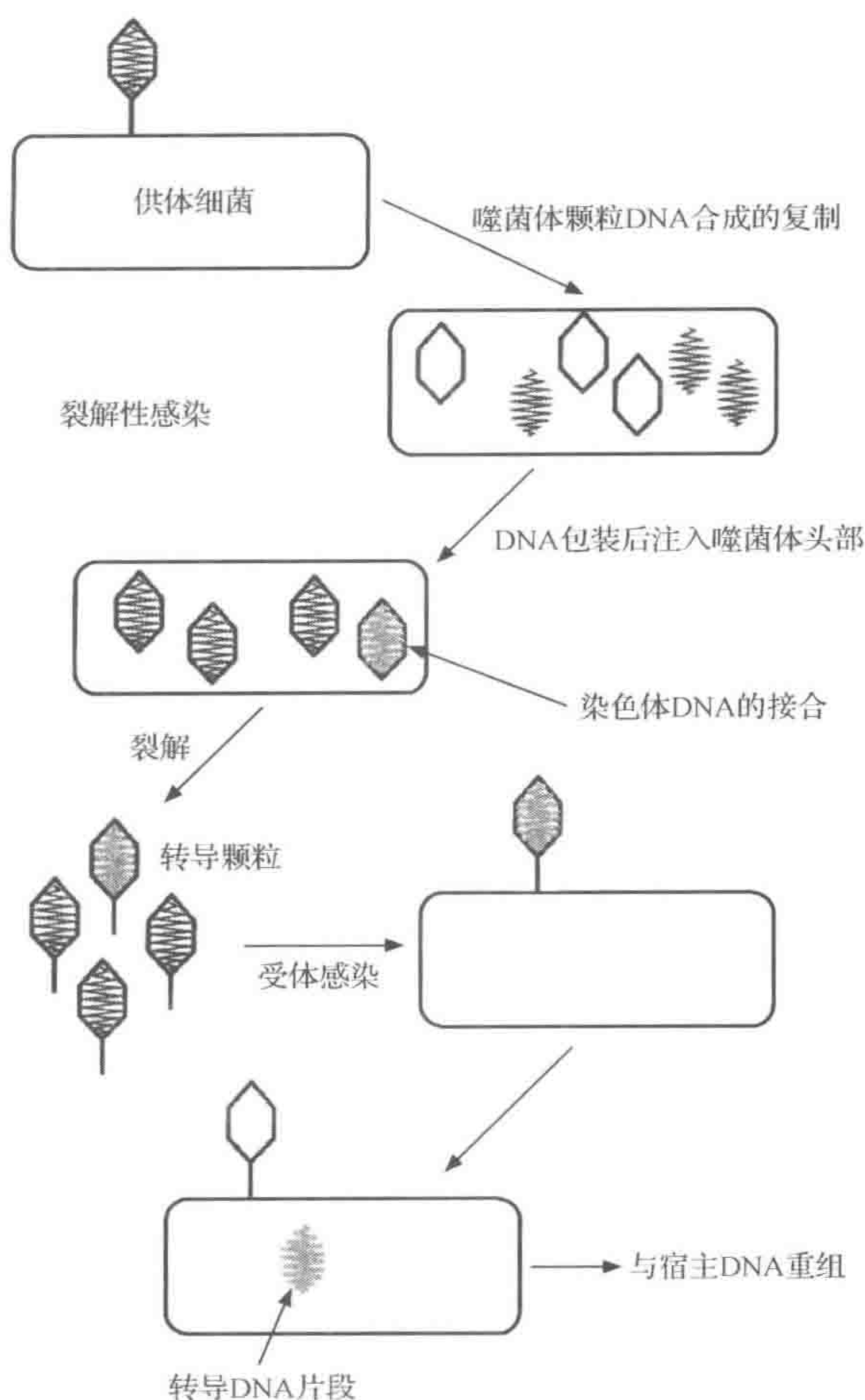


图 6.6 广义的转导

被插入到细菌 DNA,作为细菌染色体的一部分进行复制。当溶源裂解后,噬菌体进入裂解循环,通过整合的前噬菌体的各个末端间的片段之间的重组从细菌染色体上切离。如果这种重组发生在错误的位置,则相邻的细菌 DNA 区域将被整合进噬菌体 DNA。这个噬菌体的所有子代噬菌体都将含有这段细菌 DNA,一旦转导噬菌体被分离,则这个细菌片段将会以非常高的频率进行转导(在每个噬菌体颗粒中转导效率可达 100%)。因为被转移的基因限于染色体的一个很小区域,所以这种现象被称为特异性转导。这与 F′质粒的形成非常相似。与 F′质粒相比,现在可以通过创造体外重组体来较容易地向 λDNA 中加入外源基因(见第 8 章)。

另一个以相似的方式被使用的噬菌体是噬菌体 Mu(见第 7 章),该噬菌体具有可通过类似转座子的机制在染色体的多位点插入。所以,用 Mu 噬菌体创造范围更广的特异性转导噬菌体更容易,这点可用于遗传图的构建及突变诱导。

一般来说,在实验条件下,转导是很罕见的事件,一亿次噬菌体感染,只会发生一次。值得重视的是,环境中自然存在的噬菌体数量极大,这样一个罕见事件的意义也会变得非

常显著,因此,转导在水平基因转移中也具有重要作用。

6.4 重组

除了被转移的元件可以独立复制,细菌间的基因转移的一个共同特点是转移的 DNA 片段需要插入到受体染色体中,通过两个分子的断裂,交换这些片段,再重新连接。这个过程被称为重组,见第 2 章所述。重组有几种不同的形式,但在本文中主要关注的是需要同源 DNA 片段的重组,而且同源区域高度相似却不必相同,因此这种重组也被称为同源重组。

6.4.1 重组的结果

对于线性 DNA 片段(由转染、转导或接合引入)与受体细胞染色体的重组,需要两次断裂及重接(图 6.7),因此,线性片段的一部分将整合进环形染色体,代替染色体的相应片段,这个过程被称为双交叉。

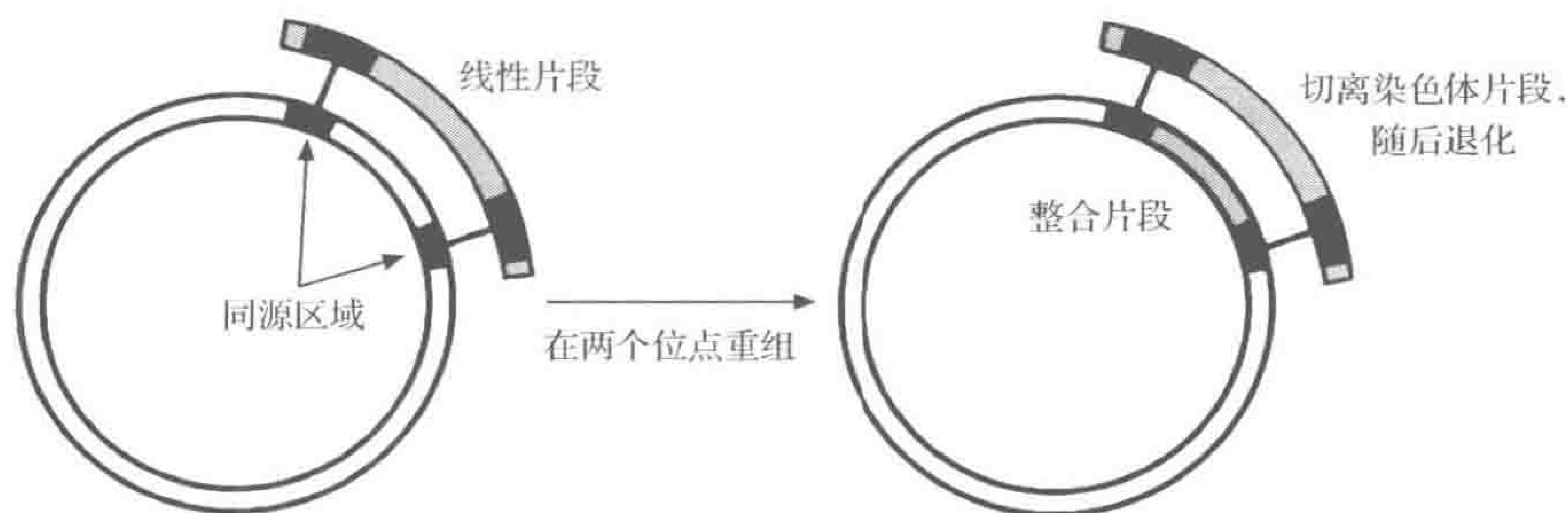


图 6.7 线性 DNA 分子和环形 DNA 分子之间的整合

两点间的重组导致环形分子的一部分被取代。所以该反应的发生至少需要两个同源区域

另外,如果两个参加重组的分子都是环形(如两个质粒或一个质粒和一个染色体),一次单一的重组事件就能够使原来的两个环融合为一体(图 6.8),这可能是一个可逆事件。被插入的质粒的两个末端间的重组导致该质粒的切离,就像前边所讲的 F 质粒的整合和 λ 噬菌体的整合切离中所涉及的那样(见第 4 章)。

在这里的讨论是假设重组发生于两个不同的分子,如果两个同源区域在同一个分子内部,即分子内重组,情况会怎么样呢?如果分子内含有像插入序列(IS,见第 7 章)一样的重复序列的两个拷贝,就会发生分子内重组。重组结果取决于两个同源区域的相对方向,如图 6.9(a)可见,如果同源区域方向相同(同向重复),它们之间的重组将导致两个分离的环形分子的产生(为了更好地理解这点,图示指出了 DNA 处于中间状态的过程:同源片段配对,从 A 点开始交换了片段,到配对区时关闭链,然而将会丢失 B 和 C 片段,直接到 D,然后回到 A)。一般来说,这种环形分子中的一个将不含有复制起始点,所以会丢失。其结果是原始 DNA 的部分片段缺失。

另外,如果同源区域的方向相反(反向重复),如图 6.9(b)所示,重组的结果不是分离出两个分子,而是反向重复序列间的片段的方向的倒位,环形分子仍然完整。

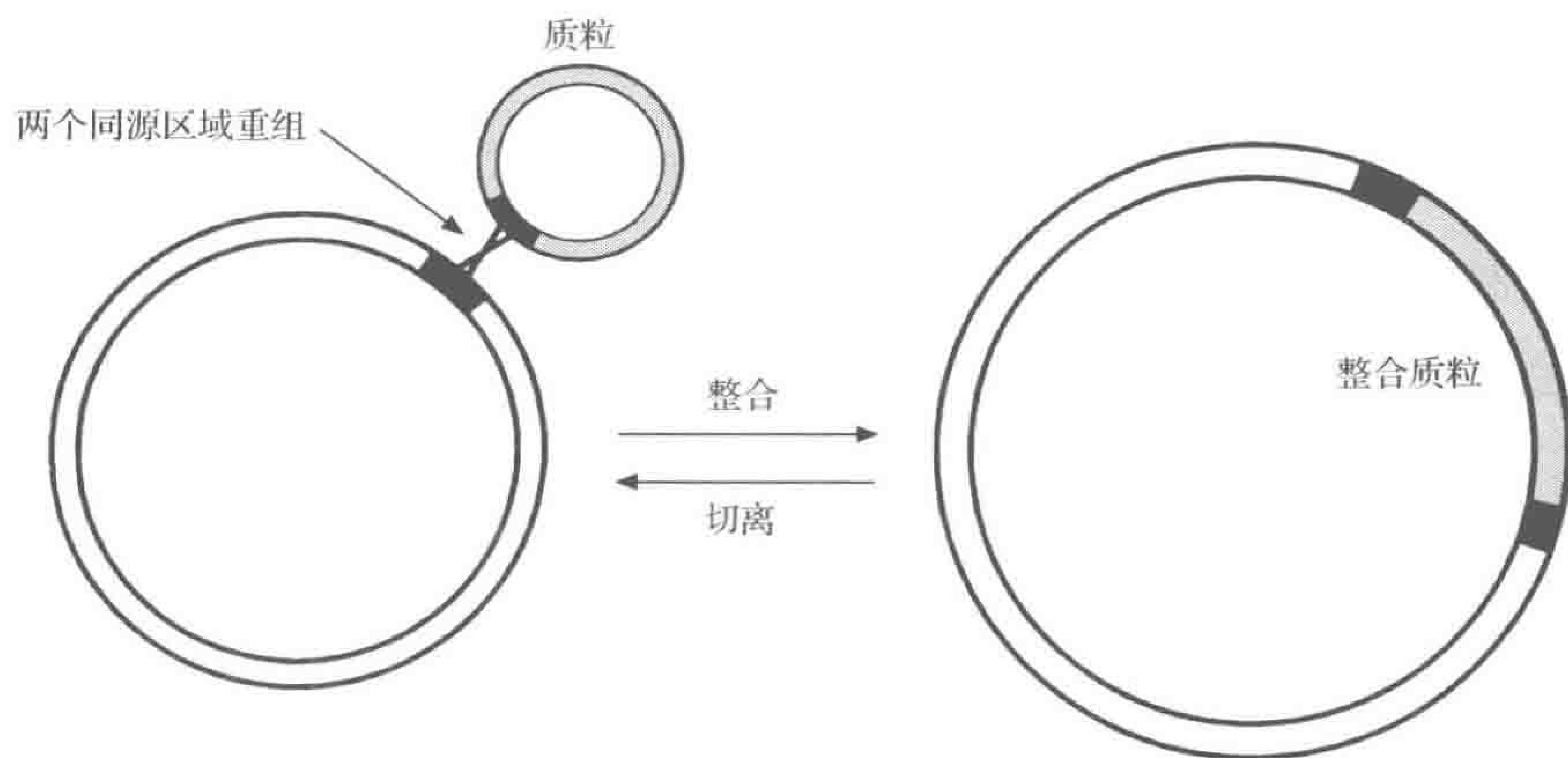
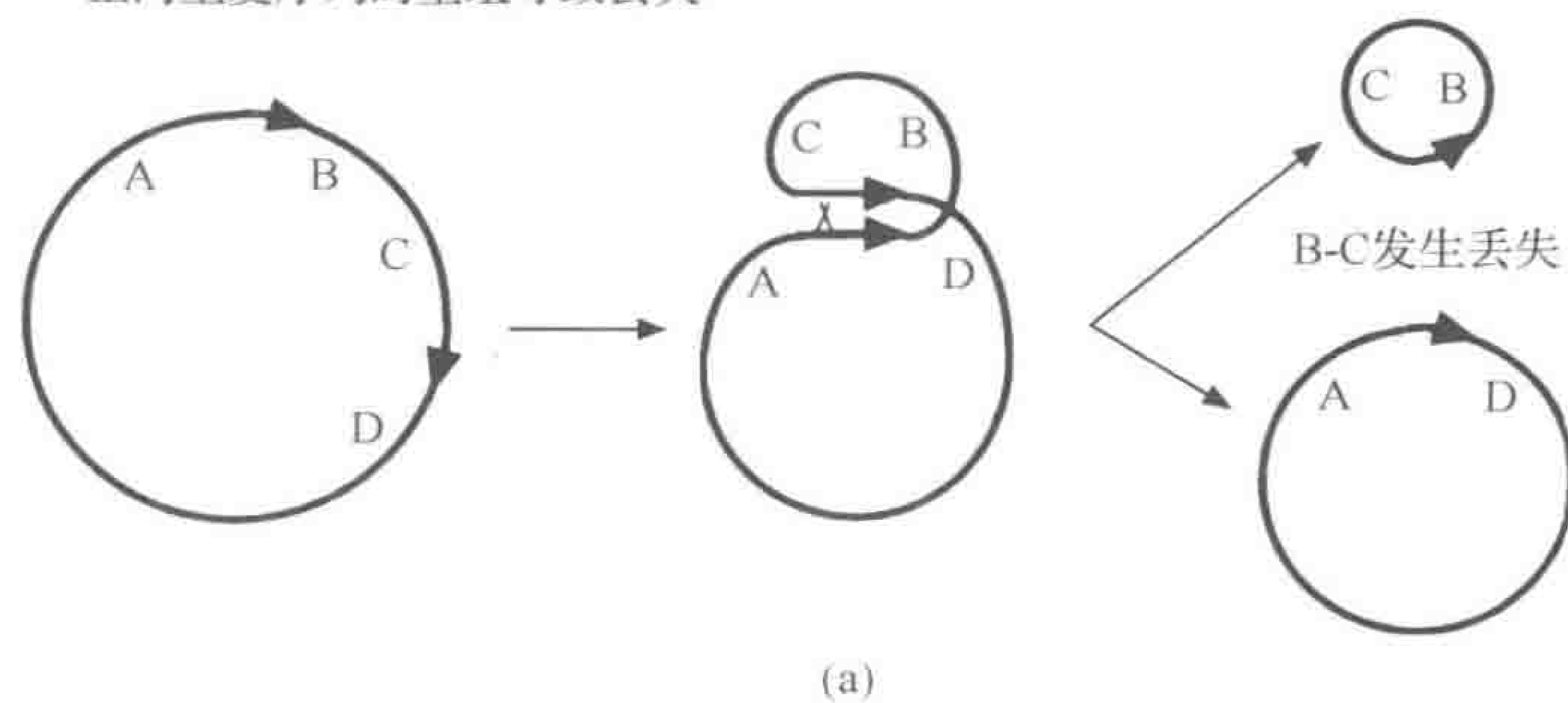


图 6.8 两个环形 DNA 分子间的重组将会导致整合的发生

单个位点的重组(单个位点的交叉)足以形成整合



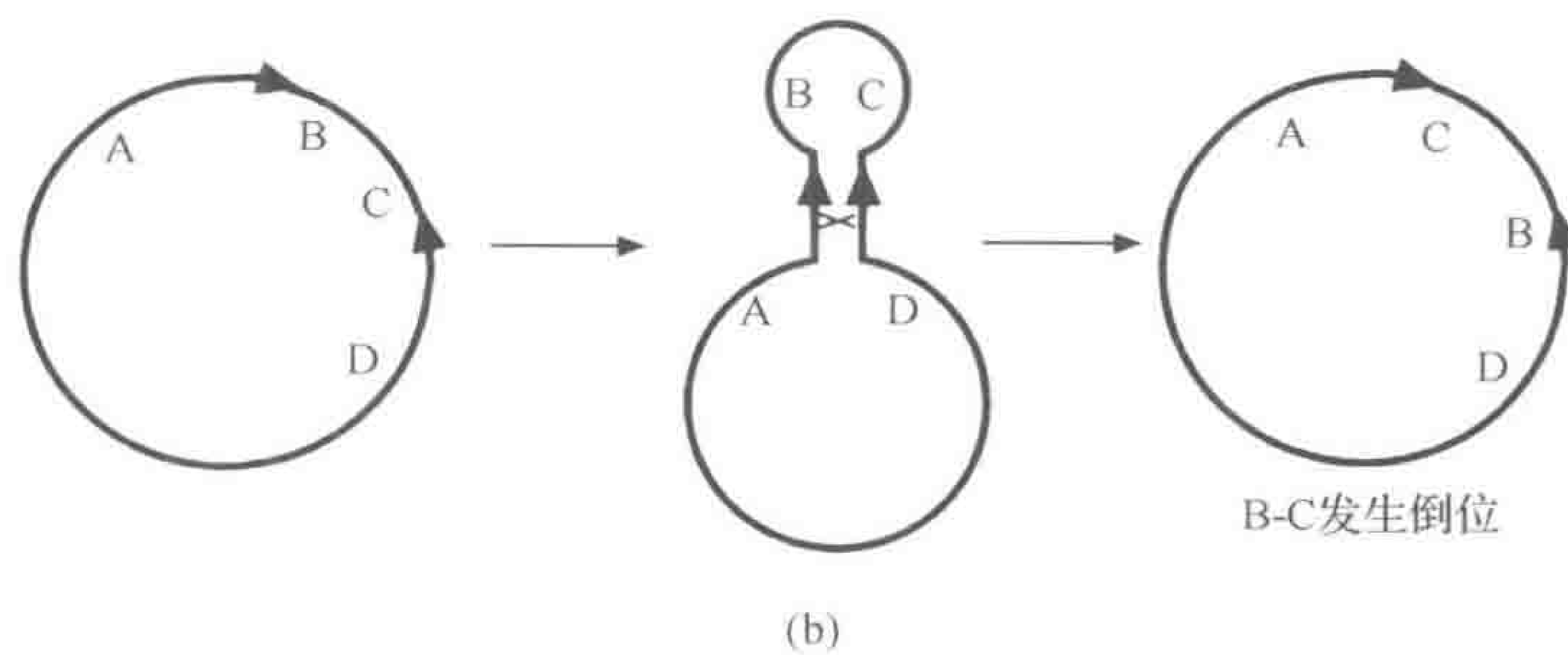


图 6.9 重复序列分子内重组

(a) 两正向片段间的重组导致片段的丢失。(b) 反向片段间的重组导致片段倒位

通过这种方式,重复元件的存在可以导致染色体区域的缺失或倒位,也可以导致质粒的重排,如第 5 章所述。例如,如果染色体中含有一个插入序列的两个拷贝,位于同一方向,而且距离相当近,则这两个 IS 序列间的重组将引起它们之间染色体区域的缺失。这可能是细菌菌株间变异产生的重要原因,如分枝杆菌属的结核杆菌,其中很多菌株的变异就是由于 IS 间重组导致的缺失。

6.4.2 位点特异性和非同源(异常)重组

DNA 分子间的重组也可以以其他方式发生，这种方式既不依赖于大范围的同源区域的存在，也不需要 RecA 的作用，这方面的例子已在本书的其他章节讲到，包括噬菌体的 λDNA 整合进染色体。这涉及噬菌体 DNA 的一定序列与染色体特异性位点之间的位点特异性重组(见第 4 章)。可移动元件(插入序列和转座子)的转座的讨论见第 7 章。

6.5 嵌合基因和染色体的可塑性

基因转移和重组不限于整个基因，也可能涉及基因的某部分。如果被转移的基因片段来自不同的菌种，这将导致嵌合基因的出现。在嵌合基因中，某些片段与其他常见片段存在根本的不同(图 6.10)。在本章前面部分提到的肺炎球菌的抗青霉素能力的产生就是一个例子。检测抗青霉素菌株的 *pbp* 基因(编码青霉素结合蛋白，该蛋白是青霉素的作用靶标)，结果发现，与青霉素敏感菌株相比，它们的某些片段截然不同，区别之大难以用简单的突变来进行解释。然而这些片段与发现于口腔的自然抗性链球菌的片段具有高度的相似性，如缓症链球菌(*Streptococcus mitis*)。从这点可以推断，*pbp* 基因的部分片段已被来自这些其他菌种的 DNA 所替代。

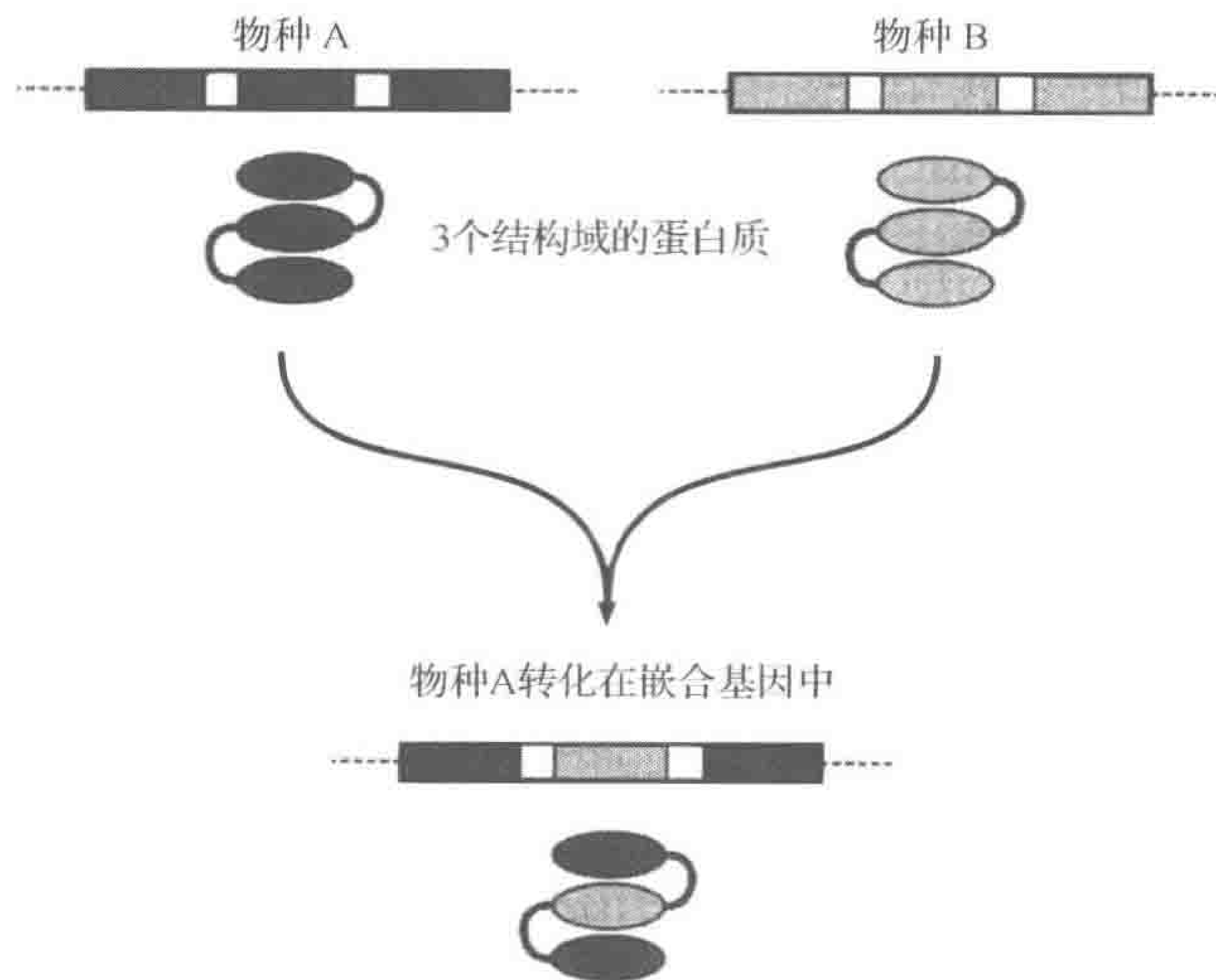


图 6.10 嵌合基因和结构域改组

图中所示的基因可以编码一个可折叠成 3 个结构域的蛋白质。嵌合基因含有来自物种 A 的两个结构域和来自物种 B 的一个结构域

基因的一部分容易被相应蛋白质结构的其他 DNA 片段取代。很多蛋白质折叠成相对独立的结构，即结构域，这些结构域由柔性环进行连接(图 6.10 和第 1 章)。在酶的活性中，每个结构域都有其各自的功能。例如，磷酸化酶(激酶)中可能有一个结构域结合 ATP，而第二个结构域则结合底物。因此，在不破坏所编码酶的全部结构的情况下，可能会混合及重排相关基因，这有时被称为结构域改组(domain shuffling)。

现在，重新定义细菌变异的概念及细菌基因组结构。简单观点即细菌(作为无性繁殖的生物体)仅通过突变的逐渐累积而出现变异是明显不充分的。水平的基因转移是普遍存在的(在大多数菌种)，不仅在同一菌种的不同菌株间，而且在种间及属间也发生，有时可跨很大的分类界限。从技术上讲，种仅通过突变产生变异，不存在水平基因转移的种为克隆，一个克隆内的所有成员源自同一个单一个体，所以，即便成员间可能有些逐渐的变异，但通过检测单一的特性(如血清型)将可准确地预测该克隆成员的其他特性。水平基因转移打破了这种关系，其结果是两个菌株可以在很多方面完全相同，但另一些方面截然不同，稍后讲分子技术在细胞分类上的应用时会重新回顾这个概念(见第9章)。

在本书第10章中，也会讲到细菌基因组结构的分析。比较基因组学表明，在某些菌种中，除了突变产生变异及从其他菌种获得DNA，还存在相当多的基因重排。携带大量基因的DNA片段出现在不同菌株或亲缘关系相近的菌株的完全不同的位置。这些通过重组(可能是有可移动元件介导，见下章)而导致的基因组重排为基因组的可塑性提供了证据，这点不但非常有利于细菌间存在变异，而且也鼓励人们对其进行研究。

7 基因组的适应性：可移动的基因和相位变化

传统观点认为，DNA 碱基的固定序列只有通过基因突变或在细胞交换遗传物质而发生基因重组时，才偶尔发生改变。现在知道，事实并非如此。实际上，遗传物质的变异是非常常见的，其结构可以发生较大规模的变化，包括插入、转位、倒位和缺失。有些变化是容易逆转的，并产生高水平的基因多样性，从而使细菌在恶劣和不断变化的环境中生存下来。有些变化是相当频繁的，且很容易被检测到，这提供了区别不同菌株的灵敏方法（见第 9 章）。

7.1　插入序列

许多遗传元件，包括一些噬菌体和质粒（见前面章节），可以被插入到细菌染色体中。然而，现在所关注的遗传元件通常不是独立存在的，只是作为染色体的一部分（或其他 DNA 分子，如质粒的一部分）。这些最简单的遗传元件被称为插入序列（IS）。正如将在后面看到的，参与基因重排的许多其他遗传元件具有和 IS 元件相同的重要特征。

7.1.1　插入序列的结构

有许多已知的插入序列。它们在大小和其他细节上存在差异，但大部分插入序列的整体结构是相似的。其中一个例子（IS*1*）如图 7.1，IS*1* 长为 768 个碱基，但许多其他插入子元件更长（通常有 1300～1500 个碱基）。这个插入序列中心区域编码一种蛋白质（称为转座酶），它是该元件从一个位置转到另一个位置所必需的，在插入序列的末端是几乎完美的反向重复（IR）序列，在 IS*1* 中，这一序列包括 23 个核苷酸。

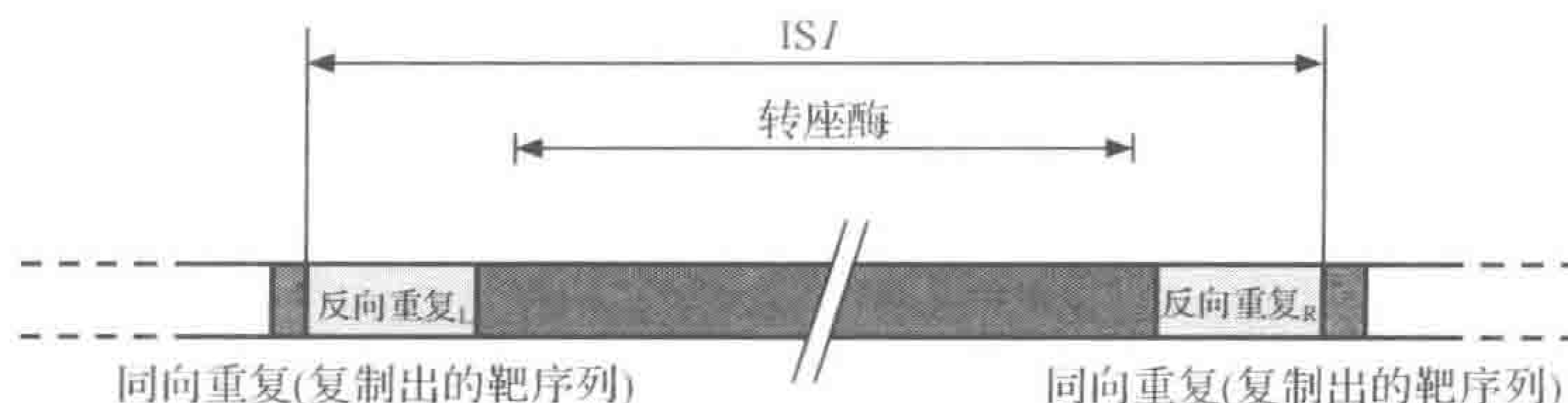


图 7.1　插入序列 IS1 的结构

必须强调的是，反向重复序列并不意味着在一条单链上反向重复，而是指在“上部”的链碱基序列从左到右的重复，在“底部”的链碱基序列从右到左的重复。也就是说，无论从

反向重复序列哪一条链的 5′端到 3′端开始阅读，其序列都是相同的。由于 DNA 序列通常只呈现两股中的一股，因此反向重复序列的 CAT 序列将显示为 ATG。

除了反向重复序列，检查包含插入序列的 DNA 区域就会发现，有一段能复制的短插入序列，但这是在同一方向重复，因此被称为同向重复序列(DR)。它不是 IS 的一部分，而是来自插入点(图 7.2)的 DNA 复制。因此，IS*1* 的不同拷贝将有不同目标序列的重复，这取决于插入点。IS*1* 的转座将产生相当长的同向重复序列(9bp)。对于其他插入序列，直接重复通常只有 2～3 个碱基。这些重复序列的存在直接关系到转座的机制，这将在本章的后面讨论。

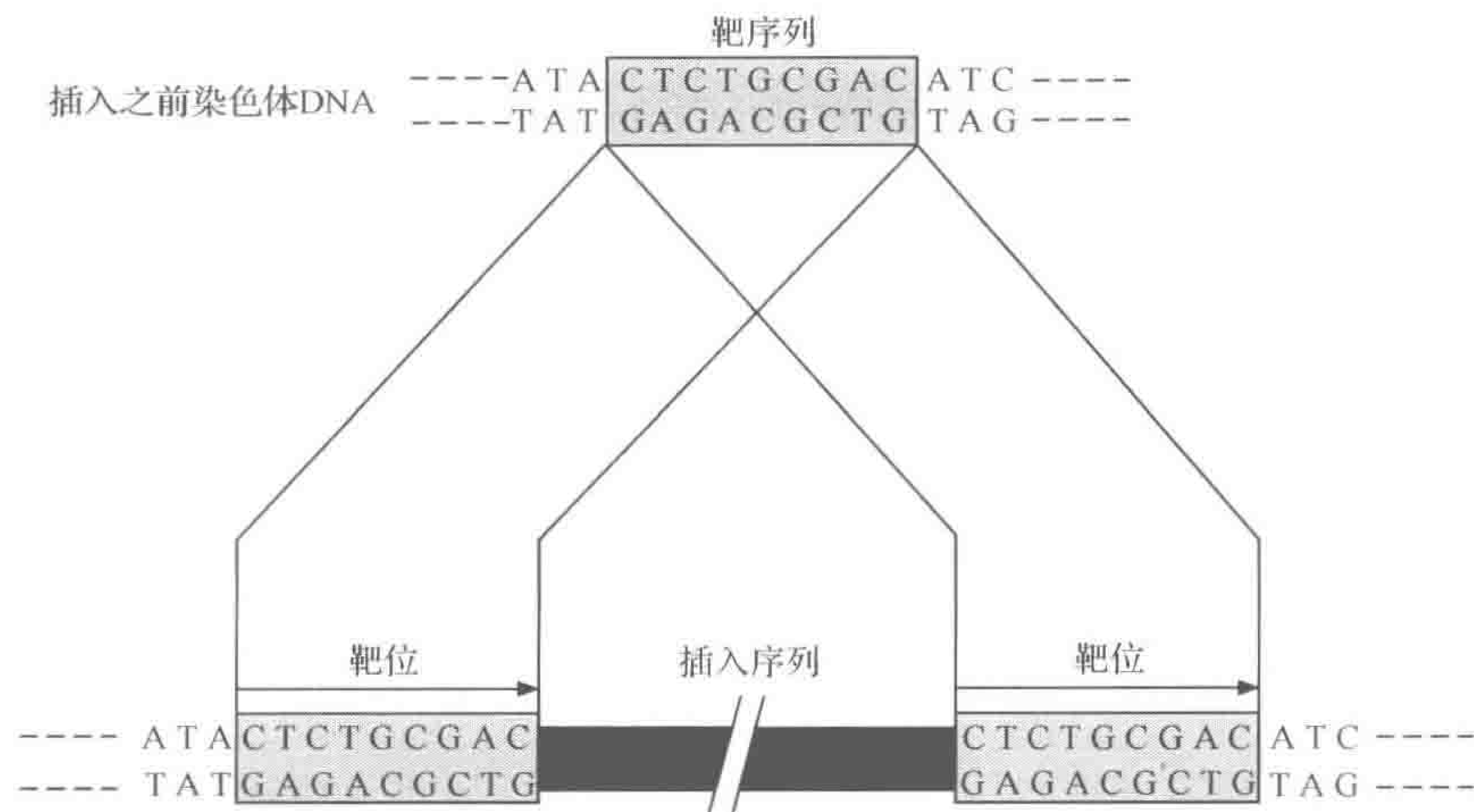


图 7.2　IS*1* 插入后靶序列的复制

IS*1* 两侧为 9 个碱基序列的同向重复，起源于靶序列的复制。示意图上方展示在 IS*1* 插入前的靶结构，下方展示 IS*1* 的插入和靶序列的复制

然而，少数元件，例如，副结核分枝杆菌(*Mycobacterium paratuberculosis*)中发现的 IS*900* 和 IS*91*(该基因被发现与大肠杆菌的毒力基因有关)，就不存在反向重复末端。这些元件以不同的方式转移，通过产生一个单链中间体，用环状复制机制得到独立的环状序列。这与第 4 章和第 5 章所描述的由单链中间体拷贝的那些噬菌体和质粒的复制方式相似。

7.1.2　插入序列的出现

尽管菌株与菌株之间经常出现某些插入元件存在与否的差异、且拷贝数也各不相同，但在绝大多数细菌的属中，都已经鉴定到了插入序列。例如，对于一株典型的实验室大肠杆菌菌株，它可能包含 IS*1* 的 6 个拷贝，分别分布在染色体中，也包含其他插入序列的大量拷贝。

利用特定的插入序列探针进行 Southern 杂交(框 2.5)，会产生一种由元件的拷贝数及其在染色体上的位置所决定的带。因为不同菌株插入序列的拷贝数及其在染色体上的位置各有不同，所以，同一菌株的不同分离子的带也是不同的。根据这一原理开发的一种名为限制性片段长度多态性(RFLP)的细菌分型方法，将在第 9 章中介绍。

IS 元件也常见于细菌质粒。例如，在第 5 章中，抗生素抗性质粒 R100 有 IS*1* 的两个拷贝和不同的插入序列 IS*10* 的两个拷贝。因为在同一质粒上两个插入序列进行重组，可能会发生基因的倒位或者插入区的基因缺失，IS 的存在是质粒不稳定性的主要原因。同样，一个染色体上存在 IS 两个拷贝也可能会发生二者之间区域的倒位或者缺失。因此，IS 元件在不同菌株之间基因结构的变异方面扮演重要的角色。

虽然 IS 元件可以通过失活或缺失基因而影响表型，但除了带有转座所需的转座酶，它们不带有任何遗传信息。因此，它们对于细菌来说很少或几乎没有任何益处。为什么细胞能够容忍它们的存在呢？达尔文适者生存的理论认为：进化压力将消除这种不利因素，因为它们的存在，会增加细胞的代谢消耗（虽然这种消耗可能很小）。最简单的答案是，这些序列本质上是寄生性的，并具有防止其被消除的策略。

对于另外一些 IS 元件，序列插入到另一位置是一个复制过程：源序列保留，而重复序列则插入到新的位点。因此，细胞内的拷贝数趋于增加。当拷贝数达到一定数量之后，这个过程则被抑制，以防止细胞不堪重负生长缓慢，甚至死亡（从而消除这种“寄生物”）。但是，如果细胞失去一个或两个拷贝之后，由于其数量不足以产生这种抑制作用，就会导致新一轮复制和插入。因此，像任何适应性良好的寄生物一样，虽然插入序列在宿主体内“定植”，但却尽量避免对宿主产生足够伤害，以防影响到寄生物自身。

7.2 转座子

当抗性质粒第一次被发现时，对于这么多不同抗生素抗性基因如何被携带，特别是这些抗性基因是如何在质粒中进行组合的（反过来讲，不同质粒是怎样携带相关抗性基因的）有很多猜测。假定一个基础质粒能独立地进行复制，但是没有携带任何其他信息，以某种方式从抗性宿主菌株染色体上获取了一个抗性基因。转移到一个敏感性菌株后，产生选择性优势，因此间接的选择这个具有优势的“新”质粒。当该质粒从一个生物体转移到另一个生物体时，它还有机会获得额外的抗性基因，从而产生了含有不同抗性基因组合的质粒家族。

由于这种模式意味着不相关的质粒可以独立地获得相同基因，这可以解释某些抗性基因的广泛分布，特别是 β-内酰胺酶（一种可以破坏青霉素，从而使其带有青霉素抗性的一种酶）。这种特殊的 TEM β-内酰胺酶，在肠杆菌属细菌的质粒中是最常见的，也存在于其他菌属的许多成员中。

TEM β-内酰胺酶普遍存在的原因，是由于发现基因可以从一个质粒转移（转座）到另一个。通过接合实验的图解所示（图 7.3）。一株包含两种不同质粒的大肠杆菌，其中一种质粒具有氨苄青霉素抗性基因而另一种质粒具有卡那霉素抗性基因，以此菌株作为供体，另一株对两种抗生素都敏感但具有萘啶酸抗性的大肠杆菌作为受体。将这两种细菌的混合培养物接种于同时含有氨苄青霉素、卡那霉素、和萘啶酸的培养基上，可以筛选得到从供体获得另外两种抗性的受体大肠杆菌。结果发现两种抗性的转移率比预测的这两种质粒的独立转移率大很多。还发现抗两种抗生素的受体菌株含有携带双抗生素抗性基因的一种质粒。

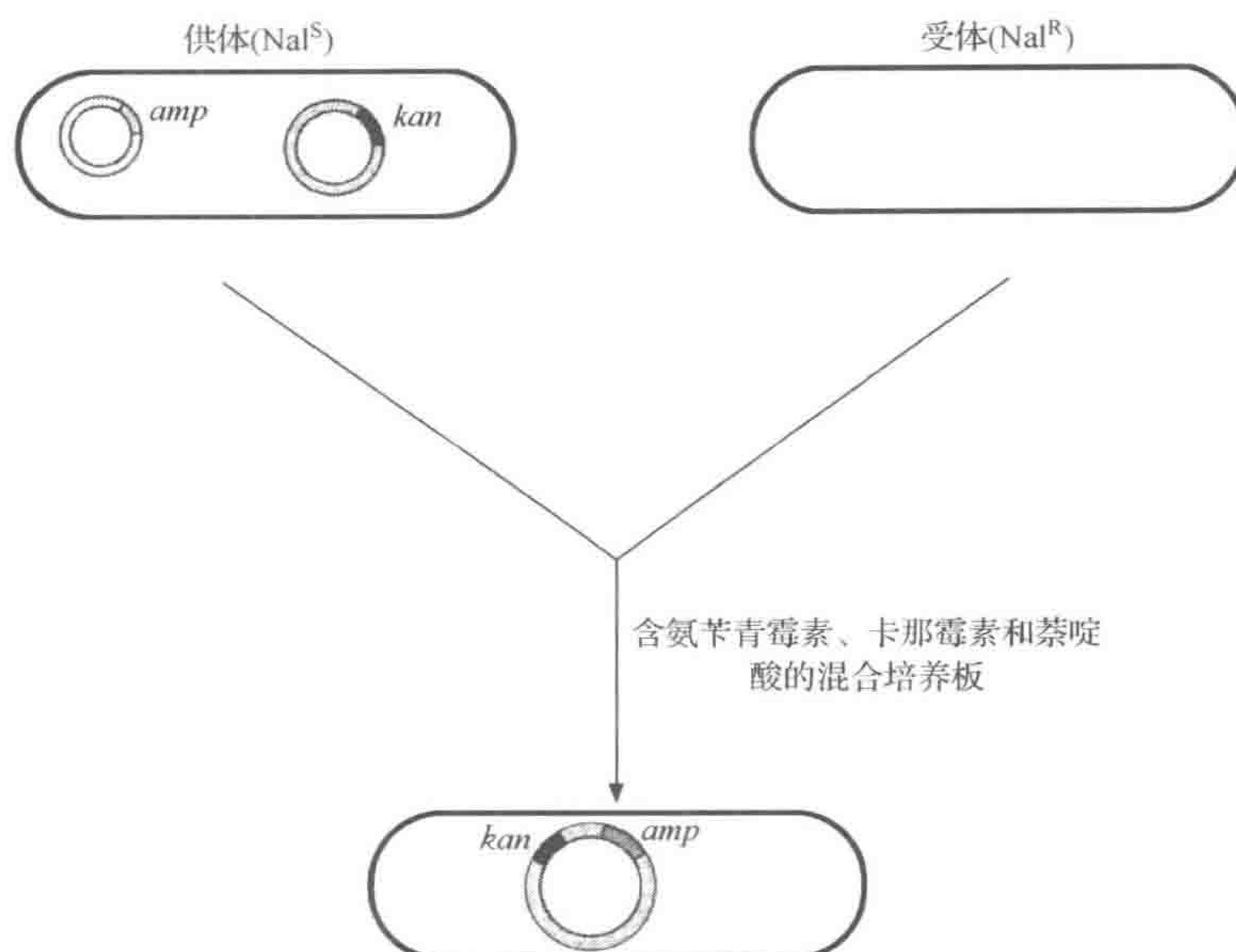


图 7.3 质粒间抗性基因的转座

供体株有两个质粒，分别携带氨苄青霉素和卡那霉素抗性基因（*amp* 和 *kan*），和一萘啶酸抗性受体结合，用含这三种抗生素的选择培养基筛选，得到的克隆质粒上同时携带 *amp* 和 *kan* 基因，这是因为 *amp* 基因转座到另一个质粒上

因为它在重组缺陷（*recA* 基因）株同样出现，所以认为这种效应并不是由于两种质粒的普通重组。从其他证据可以推断，氨苄青霉素抗性基因已经从一个质粒转移（转座）到另一个质粒。术语转座子被创造出来说明具有这样行为能力的一个元件，也就是一个携带有额外与转座不相关的可移动的遗传元件。

这种抗性基因的移动，不仅可以发生在两个质粒之间，而且可以从质粒到染色体，或者从染色体到质粒。因此，这就部分解释了为什么抗性质粒及携带非抗性基因的质粒的进化如此快。虽然这一讨论集中在抗生素耐药性上，但其他来源于质粒的基因偶尔也是可以转移的。在细菌质粒研究的文献中，抗生素耐药性研究占据主导地位有两个原因：抗生素的广泛使用提供了一个显著的选择压力而促进其发展和传播，抗生素的抗性是一个非常方便的遗传标记，从而使其比其他类型的基因更容易研究。

7.2.1 转座子的结构

一个简单的转座子 Tn*3*，其结构如图 7.4 所示，它大约由 5000 个碱基组成，两端各有一个 38bp 的反向重复序列。因此，它与插入序列是类似的，不同之处是转座子可携带一个可识别的遗传标记，如上面例子中提到的氨苄青霉素抗性基因（*bla*，β-内酰胺酶基因）。Tn*3* 还编码了其他两种蛋白质：转座酶 TnpA 和一个双功能的蛋白质 TnpR，它既可以作为阻遏物也可以对转座的一个阶段起作用（将在后面更详细的解释）。如同插入序列，在转座子的两端有一个短的直接重复序列（Tn*3* 的 5bp 重复序列）。

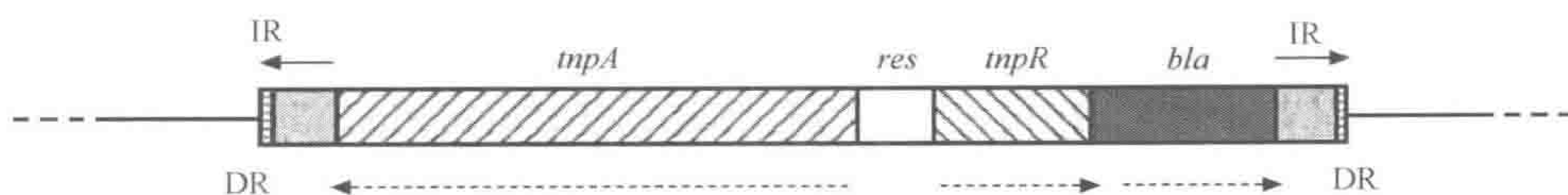


图 7.4 转座子 Tn*3* 的结构

DR，5 碱基同向重复（靶复制）；IR，38 碱基反向重复；*res*，分离位点；*tnpA*，转座酶基因；*tnpR*，解离酶基因；*bla*，β-内酰胺酶基因（氨苄抗性）

有些转座元件具有比 Tn*3* 更复杂的结构。这些复合转座子由位于一组抗性基因两侧的一个插入序列的两个拷贝所组成。例如，四环素抗性转座子 Tn*10*，它长约有 9300bp，是由位于中央部位的抗性决定基因和位于两端的同向、反向两个插入序列 IS*10*（图 7.5，另见图 5.5）组成。IS*10* 本身长约 1300bp，两端各有一个 23bp 的反向重复末端，并包含转座酶基因。

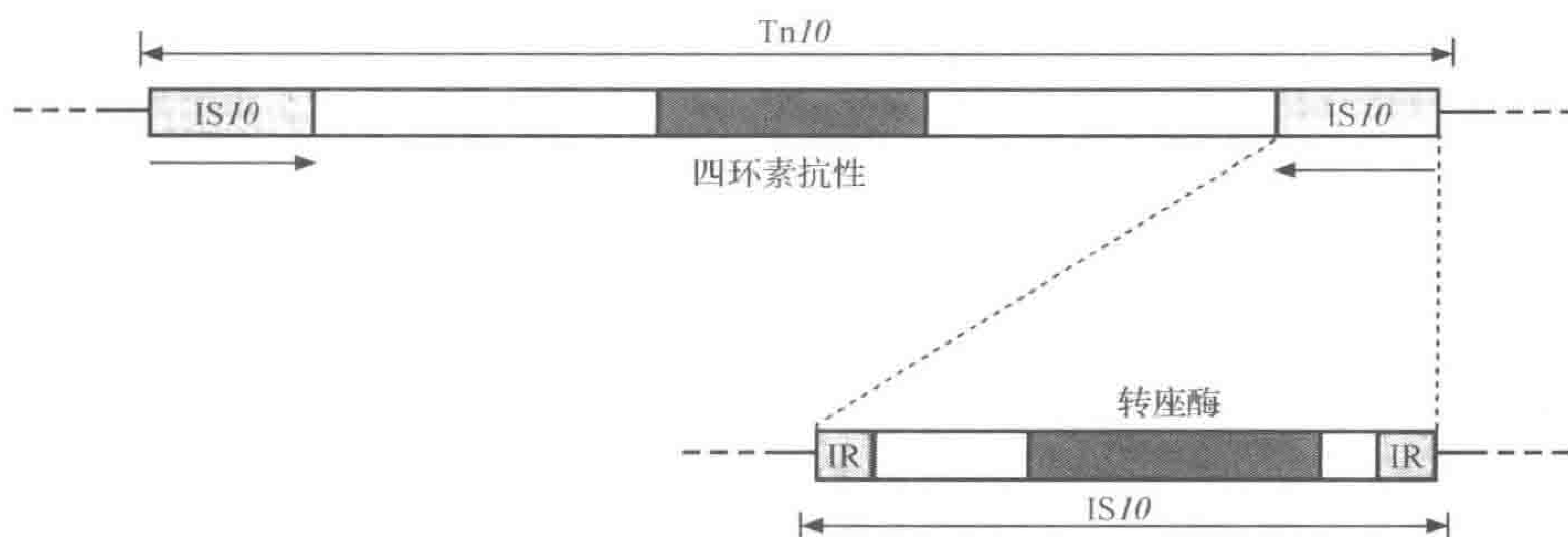
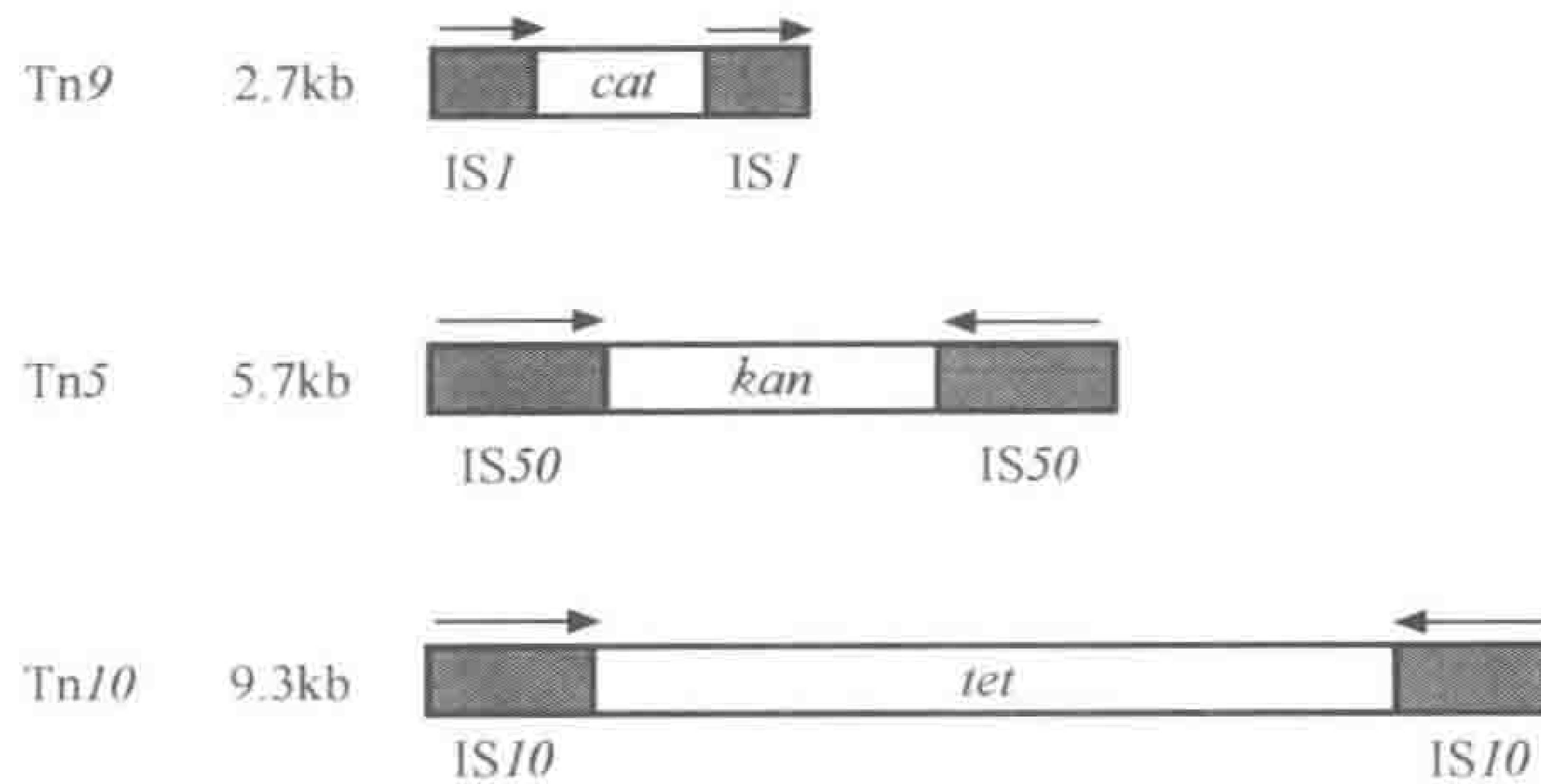


图 7.5 复合转座子 Tn*10* 结构

复合转座子侧链可能具有反向的 IS 重复序列或者是同向重复序列，例如，Tn*10* 和 Tn*5*（图 7.6）的末端均有插入序列（分别为 IS*10* 和 IS*50*）的反向重复序列，而 Tn*9* 有 IS*1* 的同向重复序列。这种复合转座子的转座可能是相当复杂的，插入序列本身可能独立地进行转座，或者整个转座子区域发生转座，而且，IS 元件之间可能会发生重组，当它们分离的时候，导致该序列缺失或倒位（见第 6 章）。

还有可能出现更为复杂的重排，例如，Tn*4* 似乎与 Tn*21* 有一定的联系，但 Tn*4* 中也包含了 Tn*3* 的全部序列。因而，Tn*4* 的氨苄青霉素抗性基因可能是完整 Tn*4* 转座子的一部分发生转座，或者通过 Tn*3* 元件发生转座。因此，转座可以在多个层面发生或相互嵌套。

复合转座子如 Tn*10*，也被称为 Ⅰ 类转座子，而 Tn*3* 转座子因两侧具有反向重复序列而不是插入序列元件，而被称为Ⅱ类转座子，或者非复合转座子。值得注意的是，这种描述仅仅是用来将它们与复合转座子相区分，并不意味着它们的结构比复合转座子简单。事实上，正如将在下一节中所见，转座子如 Tn*21*，是许多来源的组件所共同构成的。



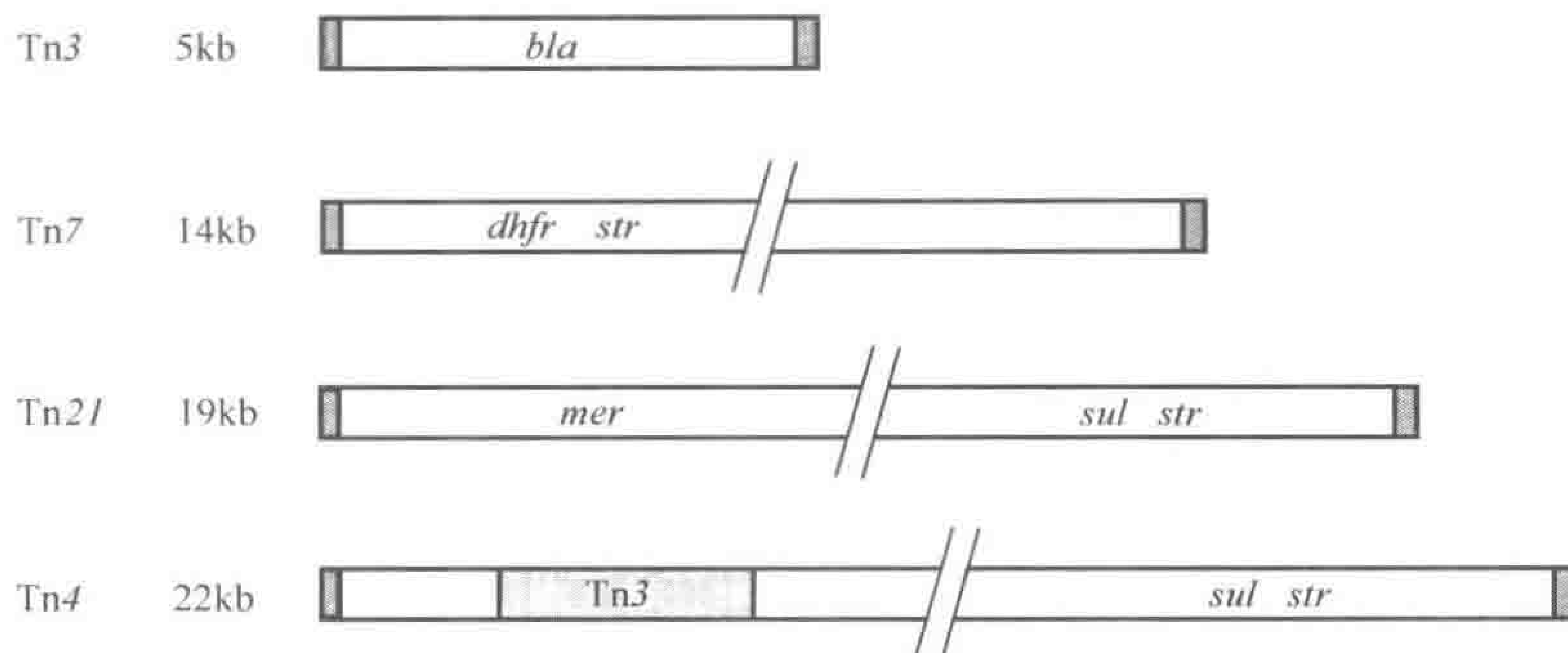


图 7.6　几种转座子结构

bla，β-内酰胺酶基因（氨苄抗性）；*cat*，氯霉素抗性基因（氯霉素转乙酰酶）；*dhfr*，二氢叶酸还原酶基因（甲氧苄啶抗性）；*kan*，卡那霉素抗性基因；*mer*，汞离子抗性基因；*str*，链霉素抗性基因；*sul*，磺胺抗性基因；*tet*，四环素抗性基因

7.2.2　整合子

通过在现有的转座子上插入额外的基因，可以形成极其复杂和大的转座子。在图7.6中已经看到这样的现象，即将一个转座子插入到另一个转座子内。但是还有其他情况存在，研究发现，许多大的转座子被证明与Tn*21*相关，但是它们携带有不同的（或多或少）抗性基因（图7.7），原因如下。

许多转座子携带称为整合子的区域，它是源自其他遗传元件的外源基因进行整合的场所。这个区域包括一个整合酶基因和一个外源基因插入的特定附着位点（*attI*）（图7.8）。一个典型的整合子在3′端也有一个保守序列（3′-CS），包含耐季铵盐（*qacE*）和磺胺类（*sul*）抗性基因。被插入的外源基因是从别处切离的环形分子，该环形分子在3′端仅包括一个不完全反向重复序列（*attC*位点）的单独基因，这就是所说的基因盒。整合酶执行着在*attI*和*attC*位点之间的位点特异性重组，这就使得在靠近*attI*位点的附近产生了基因盒整合，*attI*位点仍然保持着接受新外源基因插入的能力，使得整合子内能够建立一串由几个基因盒组成的排列。进一步剖析发现，基因盒通常不含有启动子，然而，在

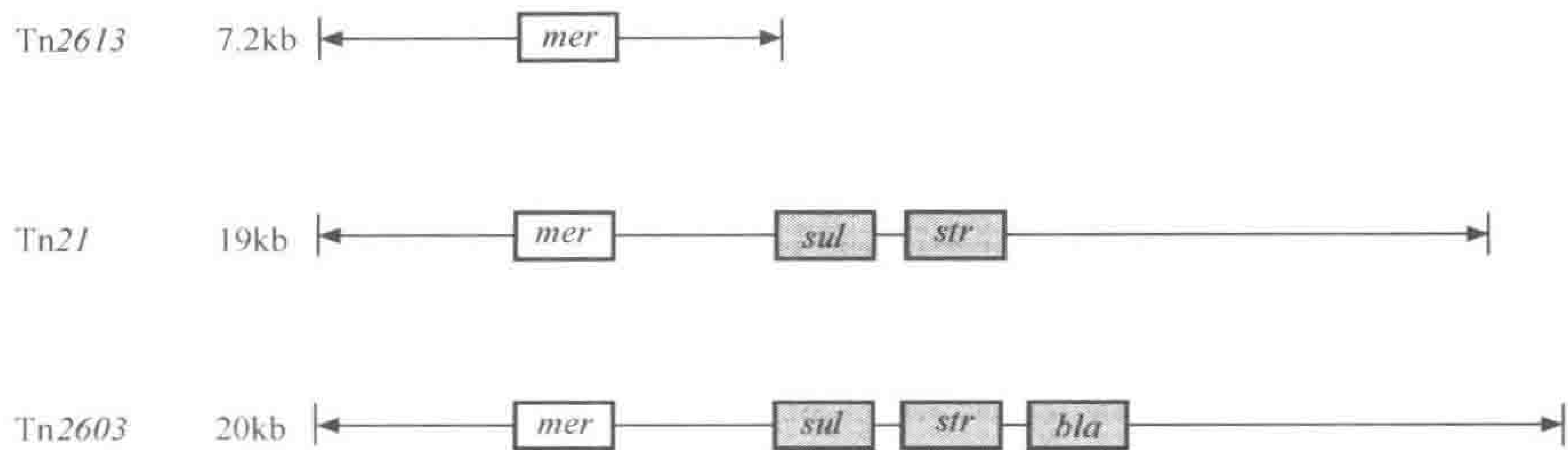


图 7.7 与 Tn21 相关的转座子家族中高度简化的代表

抗生素抗性基因标记与图 7.6 相同。末端箭头表示反向重复。与转座有关的其他基因在图中没有显示

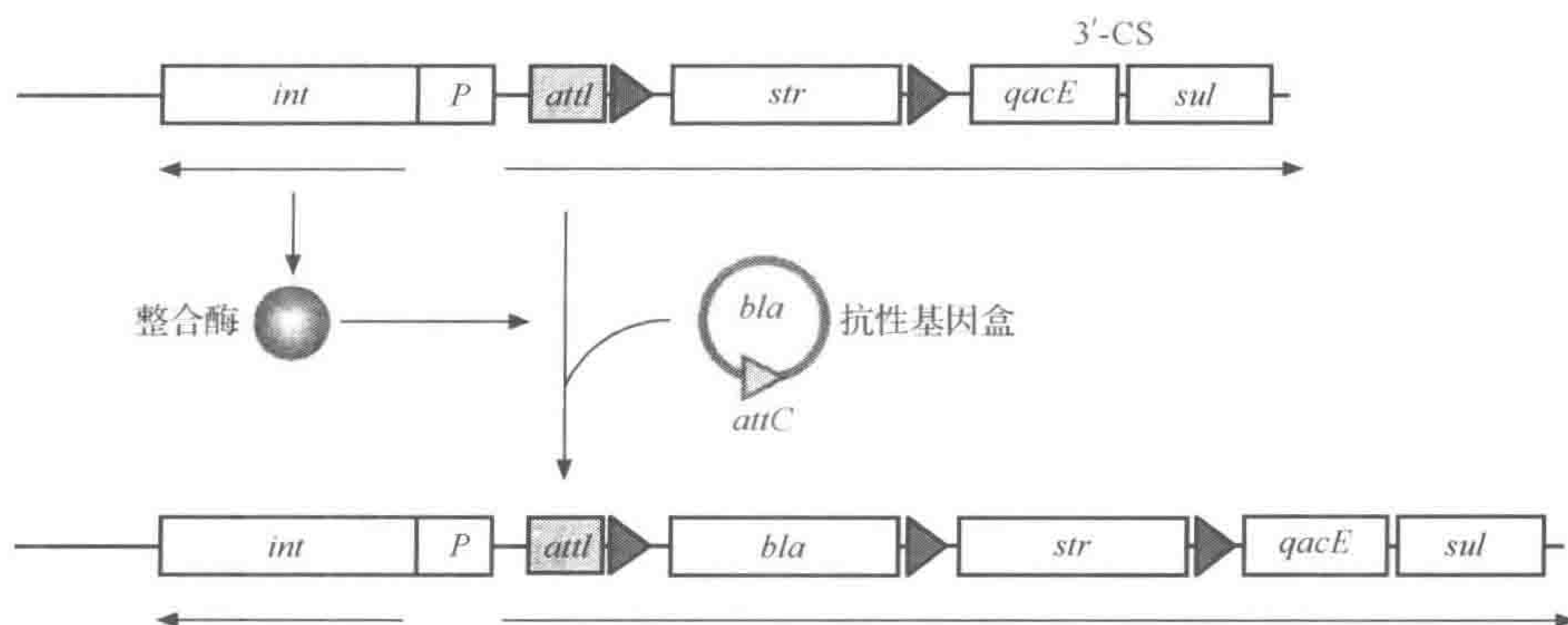


图 7.8 基因盒附加到整合区域

发生在基因盒的 *attC* 位点和整合区的 *attI* 位点间整合酶介导的位点特异性重组。3′-CS,3′端保守序列;*qacE*,季铵盐消毒剂抗性基因;*int*,整合酶基因;其他抗生素抗性基因与图 7.6 相同

插入序列的上游,整合子自身具有启动子,所以每个基因盒都是从整合子的启动子开始转录的,这样转座子就成了一个自然发生与表达载体具有类似功能的结构,关于表达载体将外源基因插入到载体靠近启动子的位置上获得外源基因的表达的内容将在第 8 章中进行介绍。

不仅在 Tn*21* 转座子家族中发现了整合子,在其他转座子和不携带有功能性转座子的质粒中,也发现了许多可被识别的不同种类的整合子。因此,整合子代表了细菌质粒进化和抗生素抗性传播的一个重要的附加机制。

虽然对整合子的认识多数来自它们在转座子和质粒进化中所起到的作用,但它们也会在其他情况中出现。许多细菌的染色体,特别是霍乱弧菌(*Vibrio cholerae*),它含有一个或多个与整合子相似的结构:具有一个整合酶、一个启动子,还有一个附着位点,后面是基因盒,但是,在这种情况下,与含有少量基因盒的典型的转座子不同的是,它们含有大量的基因盒,(在一些弧菌中超过 200 个,大约占整个基因组的 3%),它们被称为超级整合子,超级整合子的功能尚不清楚,它们所携带的许多基因的功能也不清楚,但是,它们为细胞提供了一种从其他来源获得外源基因的机制,并促进了细菌的遗传多样性,在第 9 章中,将讨论毒力岛的出现,这将为细菌细胞获得外源基因的方式提供进一步的例子。

7.2.3 ISCR 元件

正在建立一个转座子复杂结构的图片，在它们的内部可能还包含其他的转座子或插入序列，这些转座子或插入序列可能位于原来的转座子上或者是作为两个拷贝存在于一个或者更多的抗性基因的两侧（这样就可以通过形成复合转座子来转移这些基因）。它们还可能含有一个整合子，可以接纳外源基因盒的插入，可能还有另外一个复杂的因素。许多转座子被发现携带一个额外的保守区域，称之为一个共同区域（CR）。共同区域通常靠近整合子的 3′-CS 区域，紧接着是一个或多个抗性基因。这些序列缺乏基因盒相关的常见特性，并已通过另一种机制获得。CR 编码的蛋白质通常是 Orf513，它是与转座酶相关的一个蛋白，如 IS*91*。在本章前面部分提到 IS*91*，它作为插入序列的一个例子，缺乏反向重复末端，并通过滚环方式进行转座。滚环机制的一个特点就是复制不会在插入序列的末端精确的终止，而是可能延伸到相邻的基因，这些基因随后将会被整合到转座因子上，从而调动有关的基因（图 7.9）。这实际上是通过插入序列实现基因流动的另一种形式，因此，这些元件有时称作 ISCR 元件。

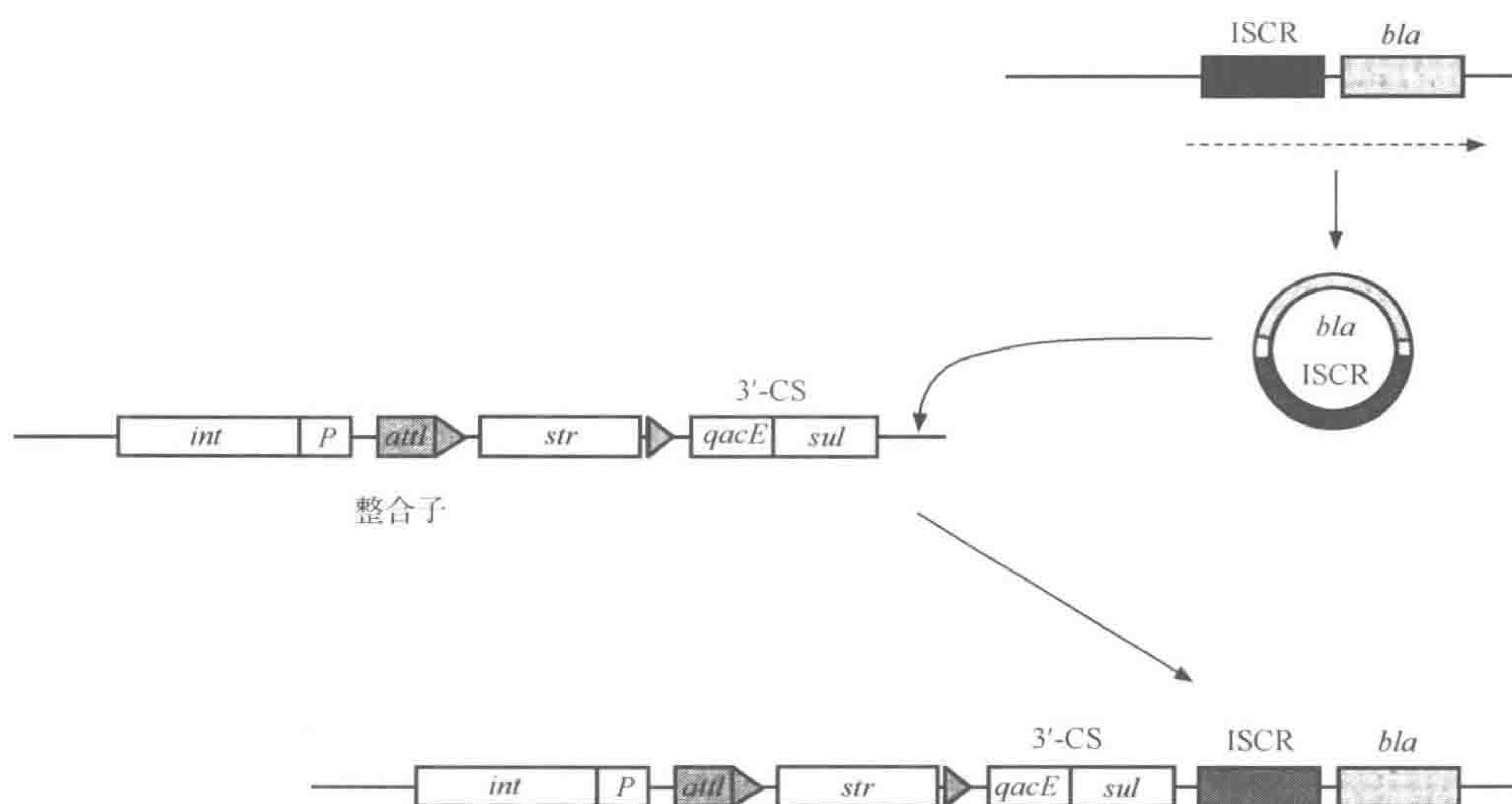


图 7.9 ISCR 元件引导的基因动员

ISCR 编码蛋白调节滚动式复制，导致含有相邻抗生素抗性基因的环形 DNA 分子的形成。这可以插入到其他位置，常见的是插入到整合区域 3′端

7.3 转座的机制

7.3.1 复制性转座

谈到转座的机制，没有必要区分插入序列和转座子，因为同样的机制适用于这两种转座元件。但是，将限定讨论那些具有反向重复序列末端的元件，对于其他的元件如 IS*91*，参见本章的前面讨论部分。

转座子如 Tn*3* 通过复制机制转座：Tn*3* 的拷贝基因插入到不同的位置(染色体上或者质粒上)，而原来的拷贝基因保留在原位置。图 7.10 列出了复制转座的过程：描述了从一个质粒(A)转移到另一个质粒(B)的转座过程。对于某些转座子，它们的靶区或多或少地显示了一些随机性。换言之，这些转座子对靶区没有特定的序列要求，而其他的转座子则对它们的靶区有特定的序列要求，一个极端的例子就是 Tn*7*，它在大肠杆菌染色体上只有一个插入位点。

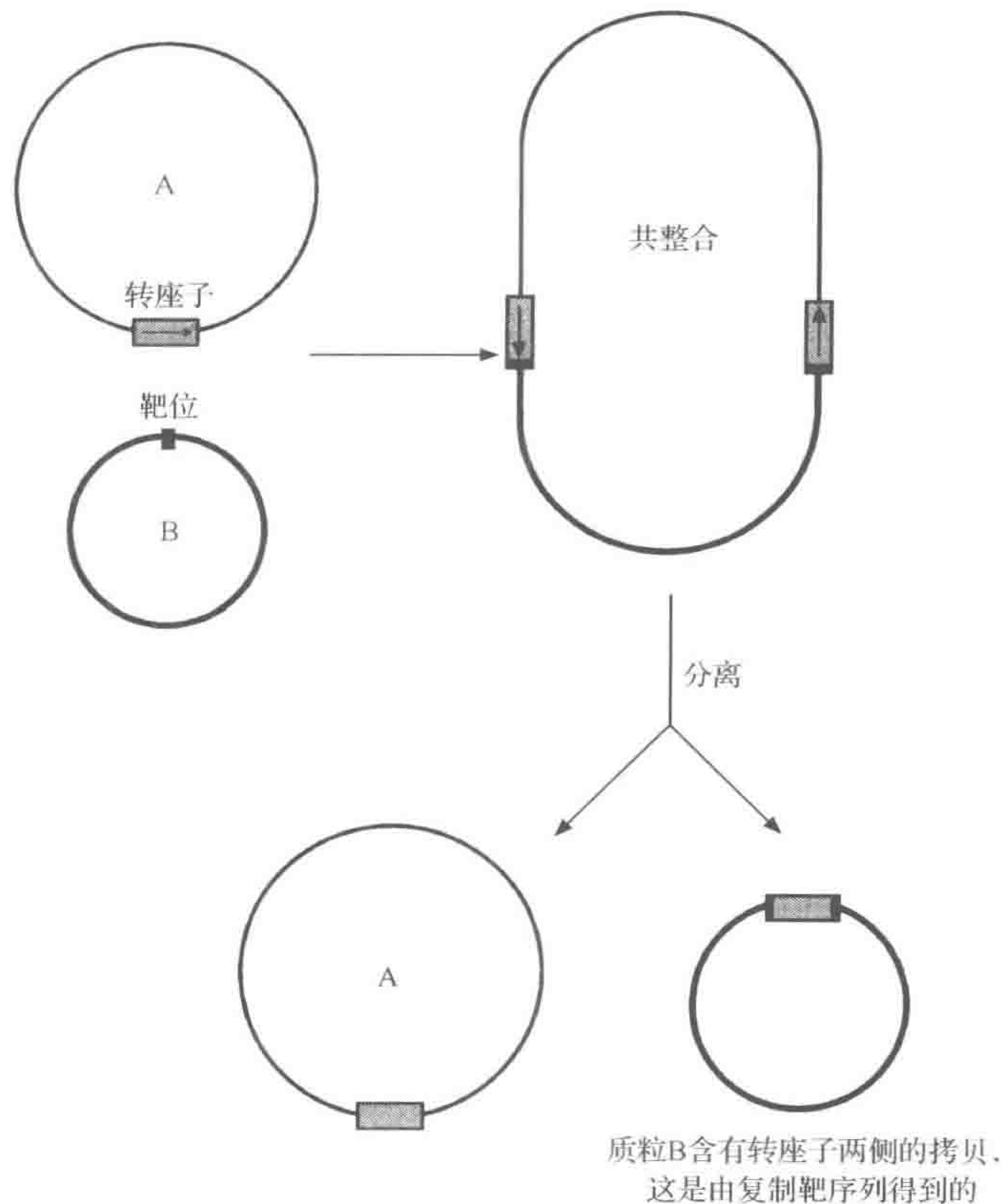


图 7.10 复制型转座子示意图

第一步显示共合体结构的形成，携带转座子的两个拷贝，对靶序列进行复制。两转座子拷贝间的重组导致分裂成两个质粒，每个质粒含有一个转座子拷贝。没有显示原始转座子拷贝两侧的同向重复

第一阶段的结果是一个被称为共联体的较大质粒的形成，它是由两个质粒以相同的方向融合在一起的全序列所组成的，但现在具有两个转座子的拷贝，转座子的新拷贝两侧是靶序列上的同向重复序列，通常有 2～4 个碱基。对于一些自然形成的质粒，可以轻易地分离到共合体分子，中间体迅速分离为两个独立的质粒。

转座子的两个拷贝方向相同(换言之，它是一个同向重复序列)，所以两个转座子在共合体分离后可以发生重组，就如在第 6 章中所作出的阐述一样，这使得两个质粒分别带有一个转座子，通过宿主重组系统进行非特异性重组，元件(或其副本)可以从质粒 A 转移到质粒 B。然而，某些转座子(包括 Tn*3*)编码它们自己的分离系统，Tn*3* 的 *tnpR* 基因编

码溶解酶(见图 7.4)，在转座子内的分离位点，它介导点特异的重组，从而保证了共合体的高效解离，而不依赖于宿主的重组活性。

复制性转座的分子基础

可以把复制性转座的分子基础看作两个时期：首先是共合体的形成，然后是共合体的分离。图 7.11 是一个共合体形成的简化模型，转座子在质粒 A 上，靶序列在第二个质粒(B)上，如图 7.8 所示。值得注意的是，线状结构是环状分子的一部分，当然这些链互相缠绕，而不是并排在一起。过程如下所示。

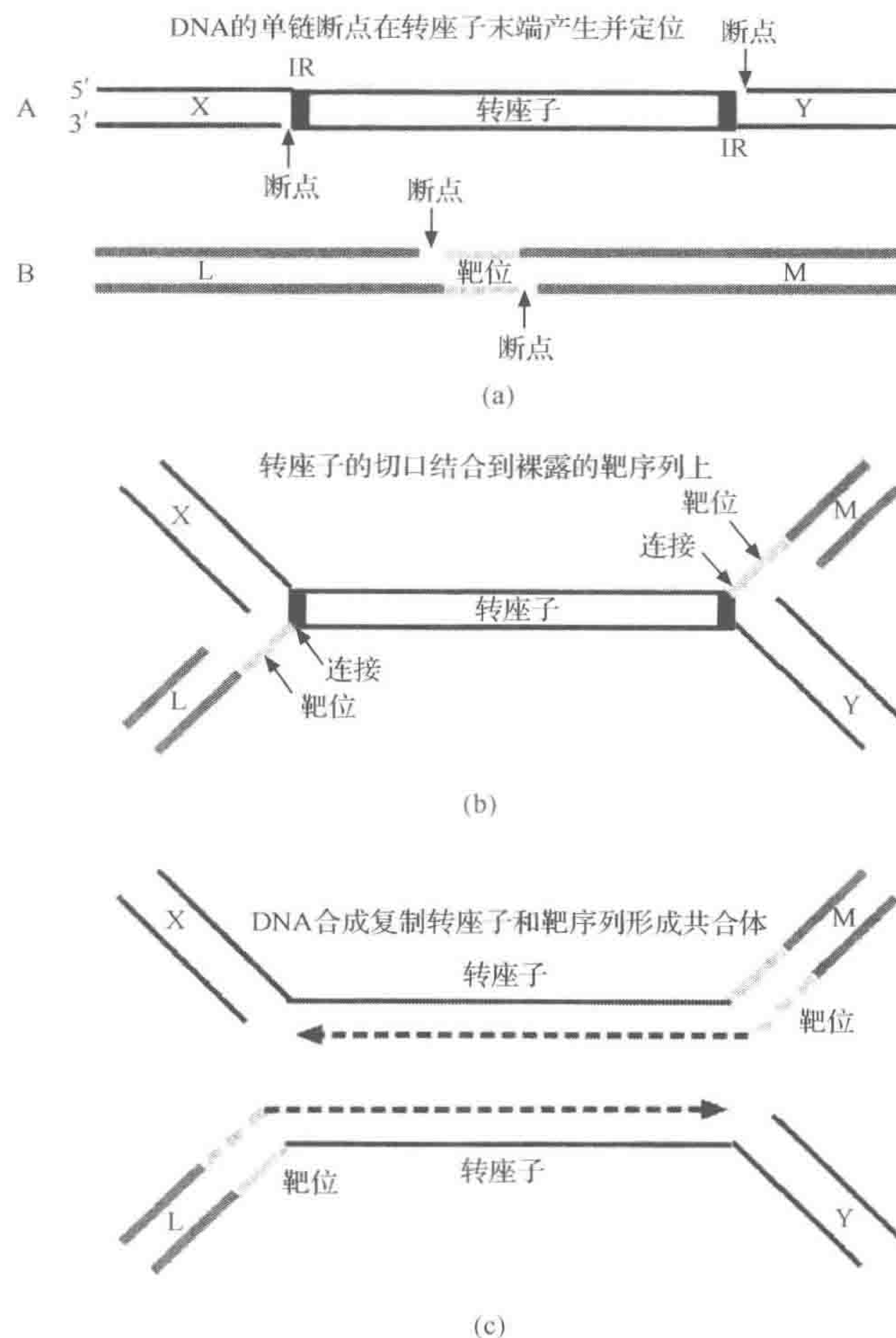


图 7.11 复制型转座子模型：共合体的形成

详见正文

(a) 在转座子的每个 3′端(在相反链内)产生 DNA 的单链断裂(切口)。受体质粒也被切出切口，位于短的靶序列两侧。缺口的交错的性质是目标序列复制的最终原因。

(b) 转座子的游离末端连接到了受体质粒切口产生的游离末端上。在这个过程中的任何阶段，转座子本身都没有作为一个独立的分子释放。

(c) 受体质粒序列的游离 3′端作为合成 DNA 链的引物，利用宿主 DNA 聚合酶合成 DNA。合成过程将通过转座子进行，两条 DNA 链分开，直至达到存在的互补链，通过 DNA 连接酶的作用，新链被连接到互补链上。该复制步骤负责转座子本身和靶序列的复制。这将产生共合体结构(见图 7.10)，即在图 7.11(c)的结构现在是一个大圈。

共合体质粒包括转座子的两个拷贝，它们是同向的(同向重复)。二者进行重组产生转座的终产物:两个质粒各含一个转座子。质粒 A 就如它开始未转录前一样，而质粒 B 上获得了一个两侧是靶序列同向重复序列的转座子拷贝。

7.3.2 非复制性(保守性)转座

不是所有的转座子都以相同的方式进行转座，特别是一些转座子和插入序列，转座时并不进行复制，这种转座方式被称为非复制性(或保守性)转座，插入序列 IS*10* 和相关的转座子 Tn*10* 都是用这种方式转座的。

非复制性转座是以一种被称作“剪切粘贴”的方式进行的，在这种模式中，在连接到受体分子前，转座子被完全从供体分子上剪切下来。因为区别主要存在于这些相关事件的相关周期内，所以这两种模式的起初表现并无太大差别。许多转座子既可以进行复制性转座，也可以进行非复制性转座，这可能是正常的。

在这一过程中，尽管对于保守性转座来说，作为供体的质粒分子减少了，但这并不意味着这种质粒会从细胞中完全丢失。由于质粒存在大量的拷贝，一部分质粒丢失可以通过复制而恢复，这将使得确定它们到底是发生了哪种类型的转座变得非常困难。

7.3.3 转座的调节

因为插入性突变的频繁发生和(在复制性转座方式中)大量的转座子或插入序列的逐渐积累，导致过度的转座可能对宿主细胞产生极严重的危害。

转座元件可以通过许多机制来控制转座的水平。前面已经提到过，Tn*3* 的 TnpR 蛋白，它不仅能够充当解离酶，而且可以作为转座基因 *tnpA* 的转录抑制物，其他的转座元件，虽然缺乏可识别的转录抑制物，但是由于它们缺乏强启动子，其转录水平较低。

一个不常见的机制在 IS*1* 和相关的 IS*3* 元件中得到阐述。在这些情况下，转座酶是由 mRNA 两个不同的阅读框翻译的，核糖体首先从某一个阅读框开始阅读 mRNA，然后在某一个确定的点退回 1 个碱基，继续沿着新的阅读框向前进行翻译(图 7.12)。因为这一核糖体读码移位的情况不经常出现，以确保由此产生的功能酶的量非常少。此外，IS*10* 体现了另外一种调节机制，它产生的反义 RNA 与编码转座酶 mRNA 的部分序列互补，可以抑制信息的翻译。

不论哪种方式，其结果都是一样的，那就是插入序列和转座子的转座都是非常少见的事件，可以从图 7.3 中很容易地得到这一结果。因为筛选得到的同时带有两种抗生素的质粒非常少。然而，如果不是通过这种显性标记，而是依赖 Southern 印迹的条带变化来寻找这种插入序列的移动，可能要检测数千个克隆才能够找到一个发生改变了的形式。

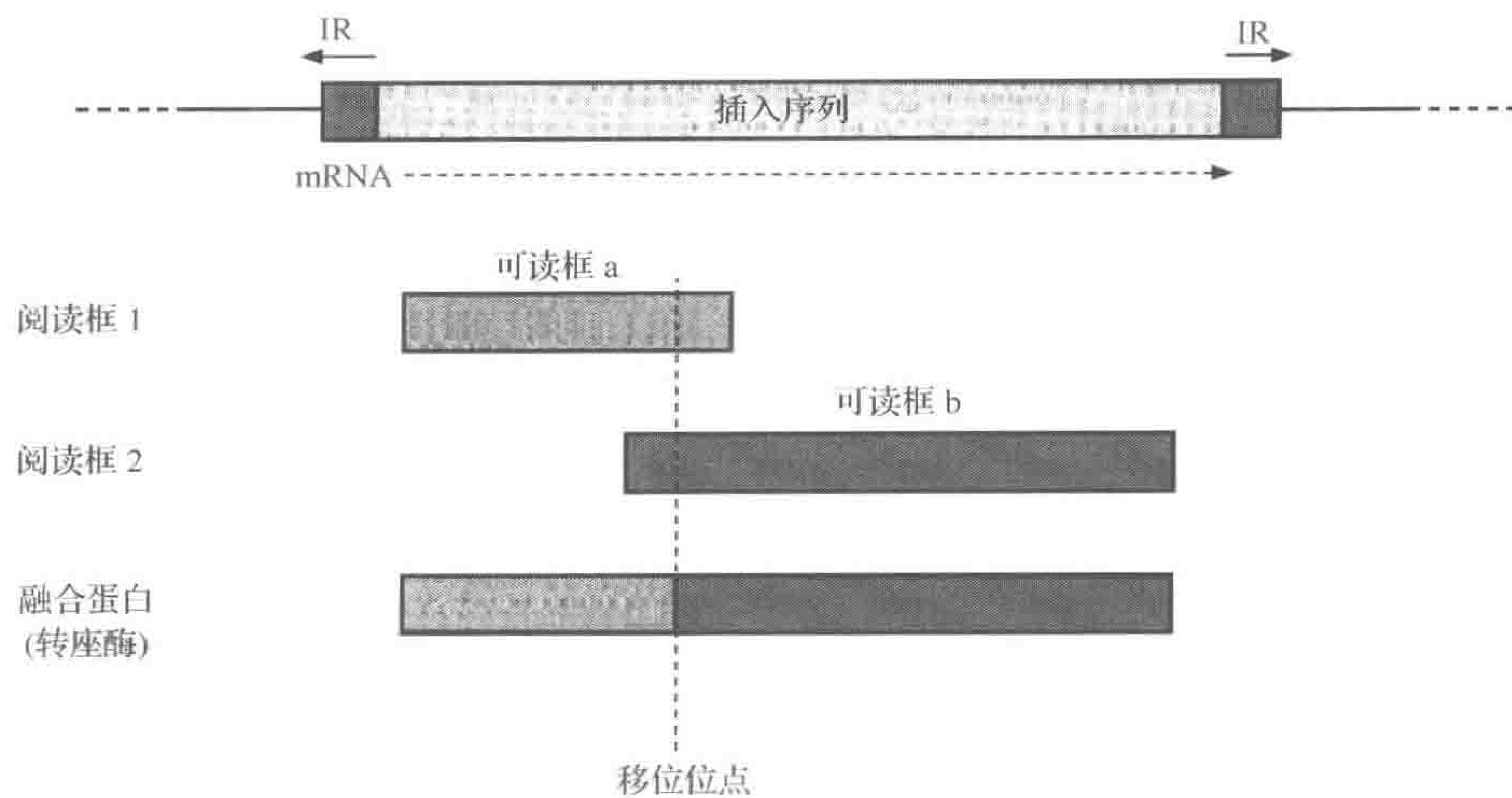


图 7.12 由核糖体移位介导的转座子调节
某些插入序列中，功能性转座酶产物要求核糖体在一中间点改变阅读框。
由于这是非经常事件，因此转座酶水平很低

7.3.4 转座元件引起的基因激活

作者得出这样一点设想，除了转座子携带的基因发生移动外，转座的唯一结果就是基因失活，这个失活的基因就是那个被插入外来元件的基因。伴随着一些转座元件的插入，相反的效应同时也会发生，也就是说，元件的插入实际上促进了邻近插入位点的基因表达，原因是一些插入序列（如 IS*10*）都含有方向向外的启动子（在图 7.13 中鉴定为 pOUT），换句话说，远离转座酶基因，朝向在染色体 DNA 的两侧的任何基因，这些基因假如方向正确，就会被 IS*10* 启动。



图 7.13 一段插入序列染色体基因的激活

7.3.5 Mu：一种转座噬菌体

Mu 是诱变物的简称，Mu 是携带有噬菌体的大肠杆菌细胞表现出异常高的突变率。诱变物噬菌体颗粒大约含有 38kb 的线状结构的 DNA，上面携带有来源于细菌 DNA 的可变末端。在注入大肠杆菌后，Mu DNA 可以随机插入到细菌 DNA 中，与转座子和插入

序列相似。这一插入过程也包括宿主基因靶位点上一个短区域的复制(5bp)。此后，通过反复的复制转座活动发生复制，这些转座基因插入到染色体上的不同位置。Mu DNA拷贝在大肠杆菌染色体的不同位置的插入引起这些基因功能丧失，因此出现较高频率的突变。

7.3.6 接合转座子

在第6章中描述了接合转座子，它们不仅可以在一个细胞中的一个位点转移到另一个位点，而且可以通过接合转座的方式转移到另一个细胞中。这些元件的转座与上面描述的转座方式是根本不同的，因为它通过切割的方式形成一个共价、闭合的环状中间物。这个结构不能复制，但是它作为底物能随之整合到不同的位点，也能通过接合转移到受体细胞。

像IS*91*这种通过环形中间体转座的元件，在本章的前半部分已经提到过。

7.4 相位变化

除了第3章描述的转录、翻译的调节机制外，许多细菌还有其他的适应机制，以使其对不同的环境做出反应。通过可逆的遗传变化，即相位变化，这些细菌被分为不同的遗传种群，图7.14描述了这种原理，这一过程设想了一个特异基因在特定状态下发生的可逆

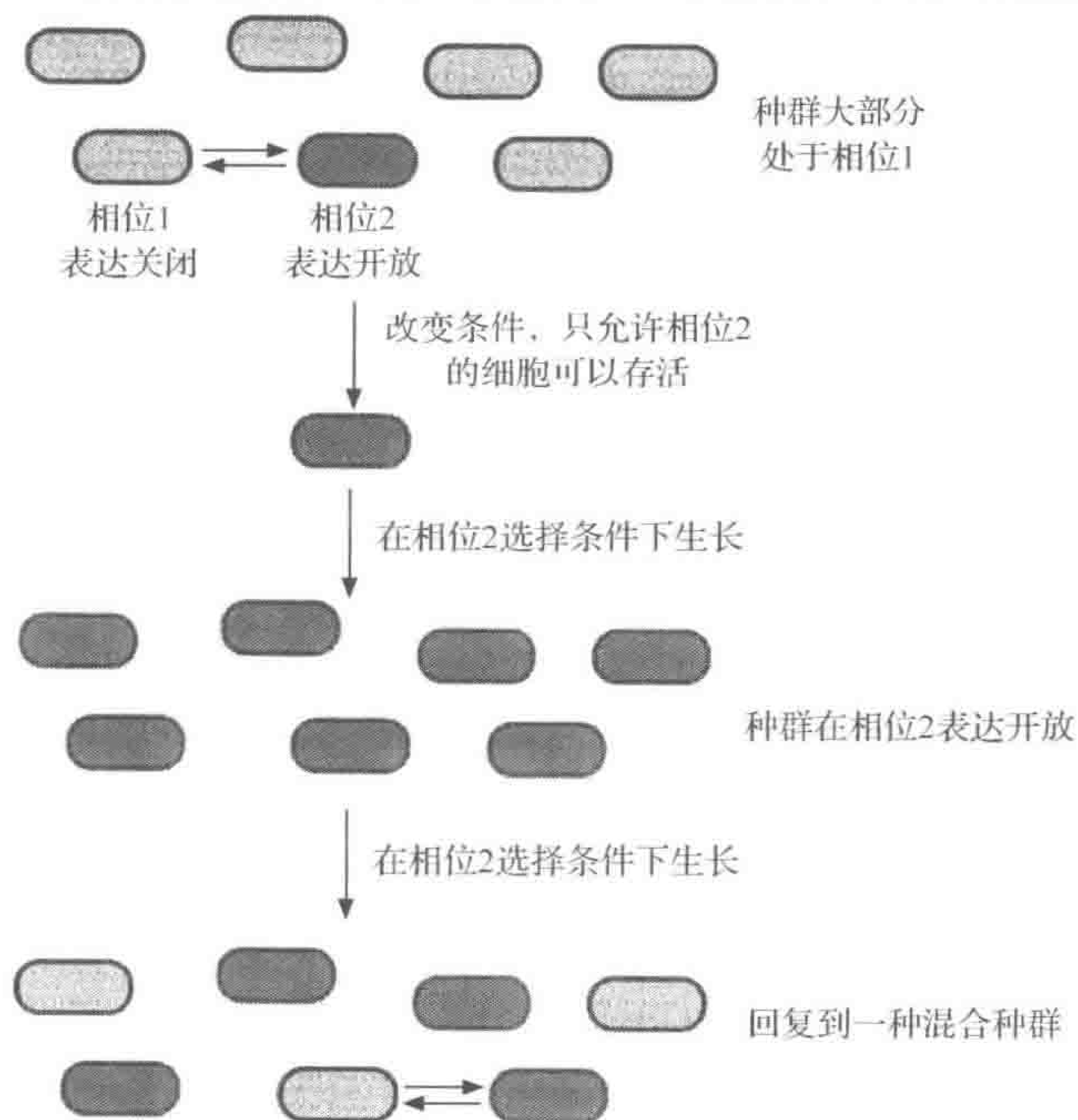


图7.14 相位变异

在最初状态，种群主要是在相位1(无特异性基因的表达)，但由于一可逆遗传变化的存在，有些细胞处于相位2(基因表达)。使细胞处于相位2细胞可以生存的环境中，将产生相位2的种群。当转换到非选择性条件时，种群将逐渐回复到混合状态

性(但是可遗传)转换。假设它会影响表面抗原的表达，相位 1，基因不表达，但是细胞转换到相位 2 的时候表达开始。最初，群体主要处于相位 1(不表达)，随着少数细胞改变状态(表达开始)，当细菌群体遭受环境变化时(如在宿主的免疫应答时)，处于相位 1 的细胞是敏感的，所有处于相位 1 阶段的细胞都被杀死，而相位 2 阶段的细胞得以存活，并大量复制，因为宿主细胞没有对相位 2 的抗原产生免疫应答，所以产生了表达该基因(相位 2)的群体。由于这一变化是可逆的，条件再次变化时，群体将会再次改变，甚至回到相位 1 占优势的阶段。一些相位变化的例子，以及所涉及的机制，将在下文中予以讨论。

可逆性的变化比常规的突变(不可逆的)更具优势。例如，病原菌既可以被宿主的免疫应答杀死，也可以被抗生素治疗所抑制。在常规突变情况下，只有一小部分病原体可能逃脱免疫反应，但是仍然对抗生素敏感，而其他一小部分对抗生素耐药的病原体可以被免疫反应杀死。然而，如果抗原变异是由于可逆性的相位变化引起的，这将是全部细胞都具有的特征(在任何时候，即使只在少数细胞中表达)，所以，能在抗生素中存活的细胞仍然能够表达不同的抗原。此外，可逆性的相位变化过程使菌群保持有潜在的遗传特性，这在一定条件下是不利的，如表达黏附和定植相关抗原，但也提供了不利于免疫应答的靶序列，在这些方面，进化选择压力可以在群体而非个体的层面发挥作用。

7.4.1　简单的 DNA 倒位介导的变化

最简单的相变系统是控制大肠杆菌Ⅰ型菌毛表达的系统。它们是位于细菌表面的，能黏附到宿主细胞上的丝状附属物。细菌在表达和非表达这种结构之间的转换频率相当高，尽管只有一个相位处于优势，但菌株的任何培养物都将包含一些相反相位的细胞，这使作者确信，如果宿主对菌毛发生免疫反应，至少有一部分细胞会得以存活，或者相反地，如果一些结构是细菌在体内启动感染所必需的，至少有一部分细菌会有这样的能力。编码菌毛结构亚单位的 *fimA* 基因的启动子，位于 *fimA* 上游一个 314bp 的 DNA 区域，其边界有 9 个碱基的反向重复序列(图 7.15)，在开放位点，启动子指向 *fimA*，并表达这一基因。在这一系统的其他两个基因，编码整合酶 FimE 和 FimB，可以作用于反向重复序列使 DNA 区域倒位。如果启动子不再指向 *fimA*，该基因则不会发生表达。这个序列的倒位须谨慎控制，在 37℃时，基因处于表达状态。然而，由于倒位机制，即使在此温度下，也总会有少量细胞不产生菌毛。

这个系统控制的结果仅是导致一个基因简单的开启或关闭，在细菌中存在着其他更复杂的相位变化，例如，大多数不同种类的沙门氏菌属(*Salmonella*)能够以相位变化的方式产生两种不同的鞭毛(H 抗原)，且具有独特的抗原性。虽然这个机制也受 DNA 区域的倒位所控制，就如大肠杆菌的 *fim* 系统一样，但倒位的结果更为复杂。H2 鞭毛蛋白的表达是由启动子所介导的，它位于 DNA 上 996bp 区域，该区域两侧为 14bp 的反向重复序列(图 7.16)。在 H2 相位，启动子驱动 *H2* 鞭毛基因的表达，同时也驱动另一种叫做 *rh1* 基因的表达，*rh1* 基因编码一个抑制因子抑制另一个鞭毛基因 *H1* 的转录。所以，在这一方向，H2 鞭毛蛋白的表达被促进，而 *H1* 基因的表达被抑制。一个位点特异的重组酶(Hin，代表 H 倒位)介导了含有启动子的区域的倒位，这个特殊位点重组酶是在倒位区编

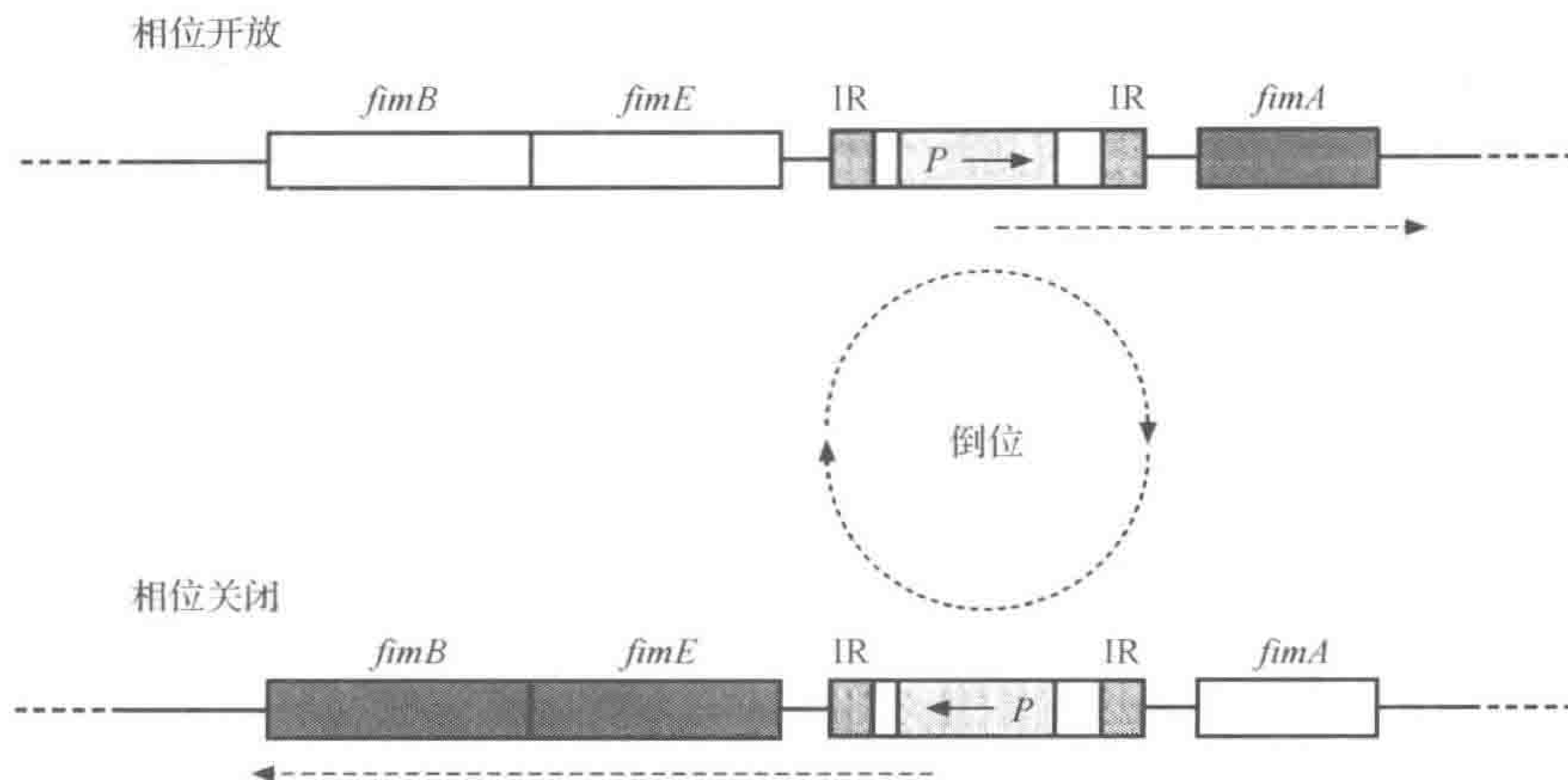


图 7.15　大肠杆菌Ⅰ型菌毛表达的相位变异。

fimA 编码菌毛结构亚基,并从位于可逆 DNA 区域的启动子表达。在第一相位,表达开放,但当控制区域将启动子面转变为错误的方向时,表达关闭。*fimB* 和 *fimE* 的产物完成倒位

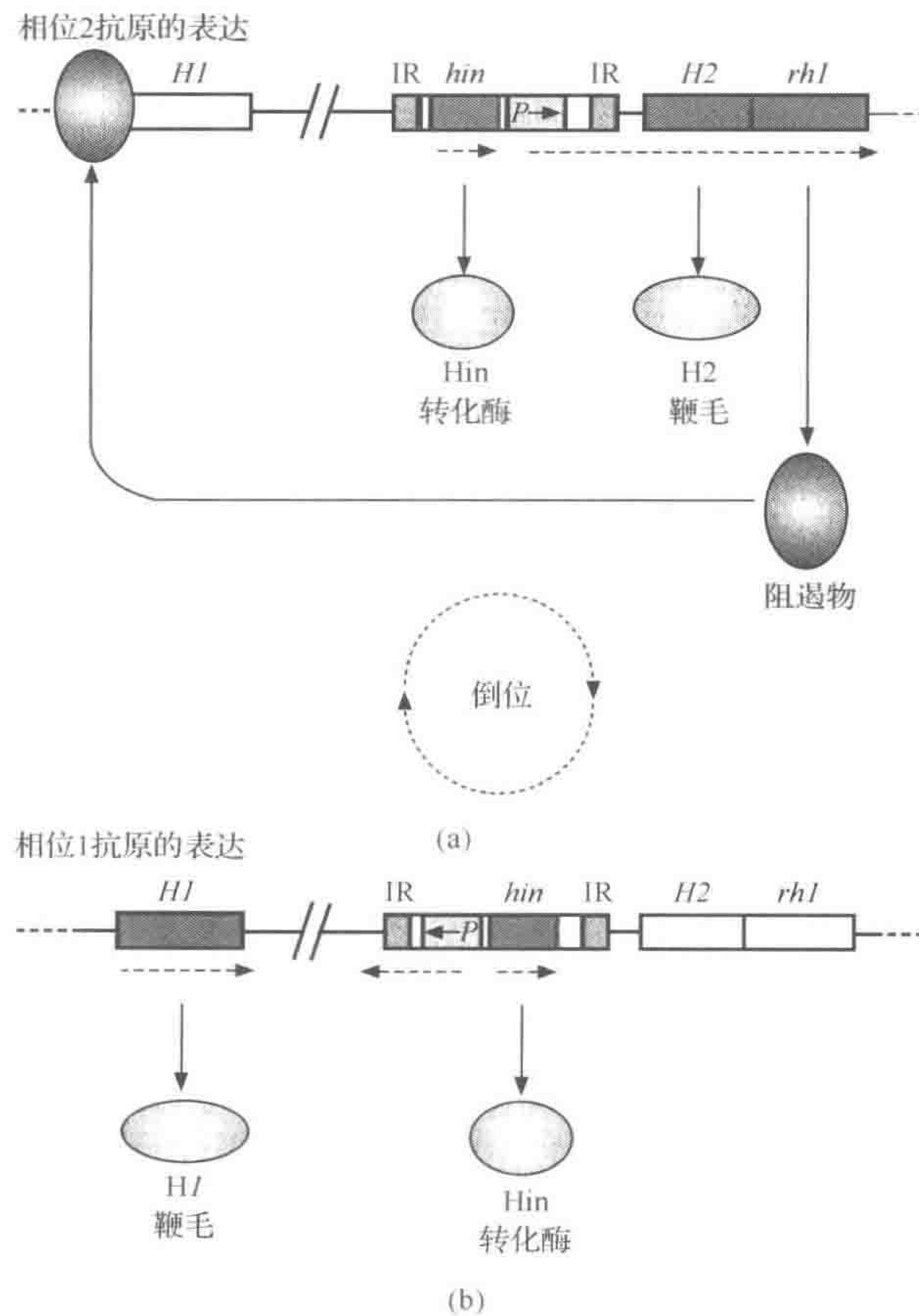


图 7.16　控制元件导致的沙门氏菌相位倒位

沙门氏菌可以表达两种类型的鞭毛(H 抗原)。(a)在相位 1,*H2* 基因表达,而 *H1* 基因被 *rh1* 基因产物抑制;(b)*H2* 和 *rh1* 控制区域开关关闭的倒位,致使相位 1 鞭毛的产生

码的，作用于反向重复序列。当倒位发生时，启动子在相反的方向，因此，*H2* 鞭毛基因不表达。由于 *rh1* 阻遏物也不表达，*H1* 鞭毛基因表达。

7.4.2 巢式 DNA 倒位介导的变化

胎儿弯曲杆菌(*Campylobacter fetus*)是家畜的一种具有重要经济意义的病原，它对宿主起防御作用的物质中有一种名为 SapA 的表层蛋白质(S 层)，虽然这使细菌可以抵抗血清的杀死，但是它也是一种抗原，也会成为抗体所攻击的目标，胎儿弯曲杆菌能处理这一问题，因为它拥有 9 种不同的 *sapA* 基因，能编码抗原性不同的蛋白质。每种细胞只表达这些基因中的一个，其他基因因为缺少启动子而被沉默。然而，DNA 重排能将启动子移动到不同 *sapA* 基因附近的位点上，允许细菌抗原表达。从图 7.17 可以看出，在两个不同方向的 *sapA* 基因之间可见启动子。环状染色体允许在每个基因 5′端的保守区域

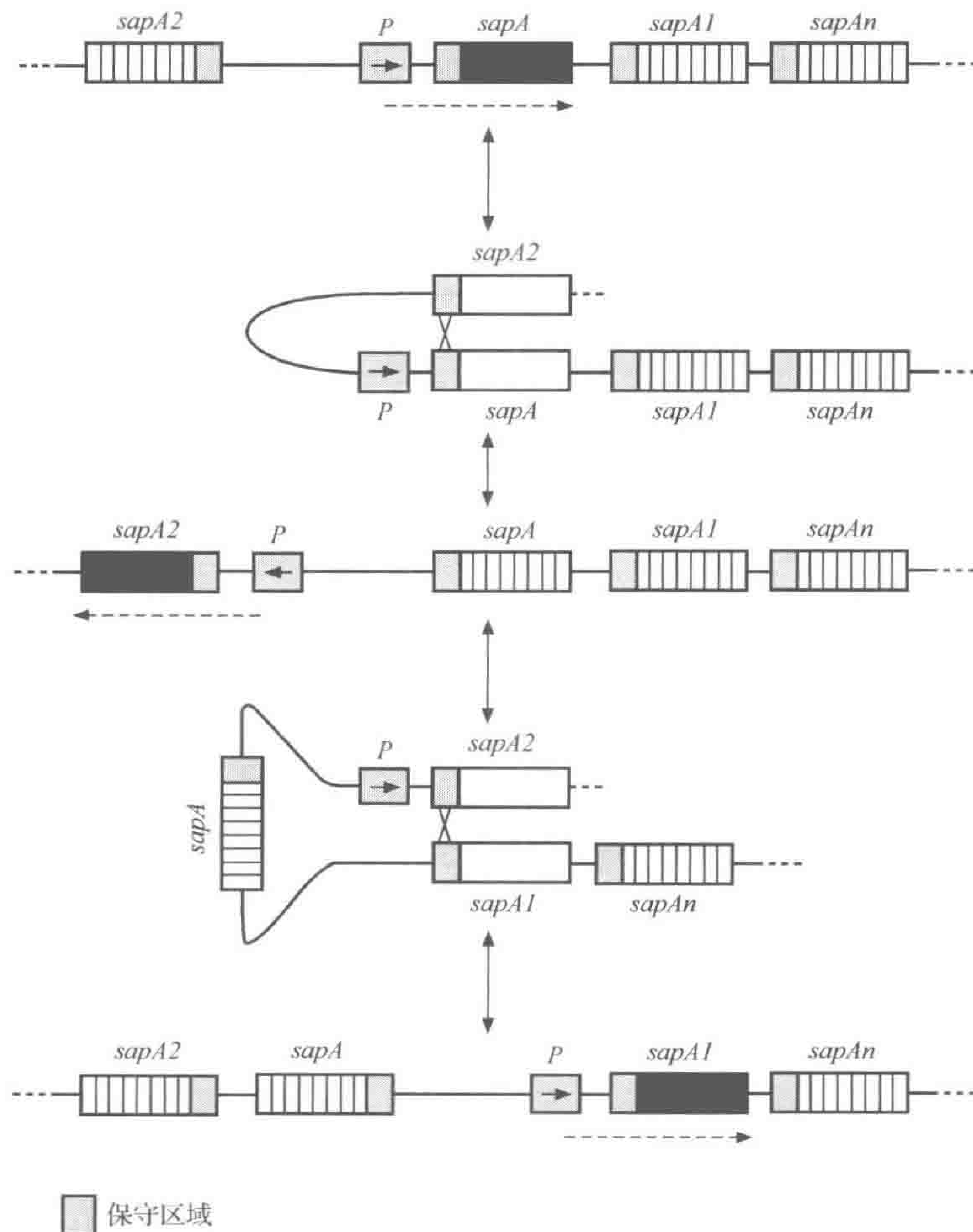


图 7.17 在胎儿弯曲杆菌(*C. fetus*)中的抗原变异

C. fetus 含有一定数量的 *sapA* 基因变种，在一时期内只有一种变种基因表达，因为其他变种基因缺少一种启动子。*sapA* 反向拷贝的 5′保守区域的重组引起控制区域的倒位，使启动子邻近于一不同的 *sapA* 拷贝

之间进行重组，从而将启动子通过包含启动子区域的倒位传递给一个不同的 *sapA* 基因。从图中可以看出，如果倒位区除了启动子外，含有一个或多个 *sapA* 基因，那么启动子可以移动到任何一个以前沉默的基因上，这个基因表达的变化导致抗原变化，使细菌能够逃避免疫应答。

7.4.3 淋球菌的抗原变异

大家熟悉的淋球菌（*Neisseria gonorrhoeae*）（革兰氏阴性菌）是淋病的病原体，它也能够通过抗原变异逃避宿主的免疫反应，这种细菌产生菌毛，这些细小的菌毛是突出在细胞表面的结构蛋白。因为在感染初期，它们能够使细菌附着于黏膜表面，菌毛是重要的毒力决定簇。淋菌的菌毛是分子质量约为 20kDa 的多肽聚合物，所有菌毛 N 端约 50 个左右的氨基酸是相同的，但是分子的其余部分（约 100 个氨基酸）差异很大。

与如上所述的胎儿弯曲杆菌的 *sapA* 系统相比，淋球菌包含一系列的菌毛基因，但是一个时期只表达一种，然而它的机制与 *sapA* 不同。图 7.18 是一个简化模型，沉默的菌毛基因被维持在一个被称为 *pilS* 的区域内，而一个 *pil* 基因在表达位点 *pilE* 内。通过用不同的 *pil* 基因替代位于表达位点的 *pil* 基因，实现基因表达的变化，从而引起抗原的变化，称为基因转换。

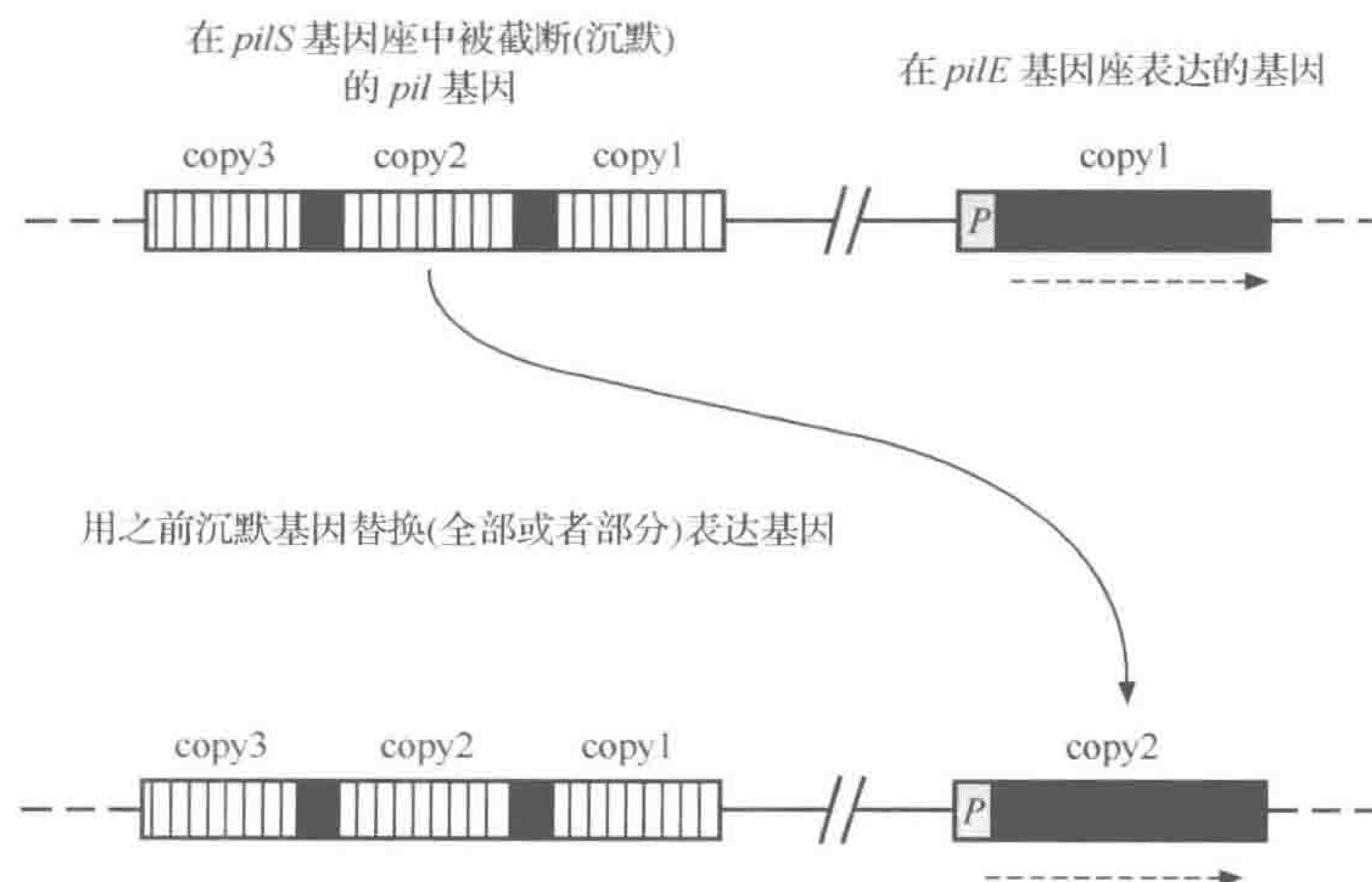


图 7.18 淋球菌菌毛的抗原变异

pilS 基因座含有多种不同的基因拷贝，都缺少启动子，因此它们都不表达。被之前沉默基因替换部分表达基因（在 *pilE*）致使抗原转变的发生。这是一个复杂系统的简化代表

由于不能发生互换的事件，基因转换不同于重组。在形式遗传学，如果用标记为 AB 的一株与另一株标记为 ab 的杂交，可能希望获得等量的两种重组体，即 Ab 和 aB。如果不能发生，或者说是 Ab 比 aB 多，这就归因于基因转换，其原因是在重组过程中，DNA 的一股在两个配对的 DNA 分子之间转换（见第 2 章），因此将获得异源 DNA 或 DNA 分子的两股不相同。在这种情况下，将获得一股是 B、另一股是 b 的区域，两股之间的错配可经过错配修复（见第 2 章），这能导致一股的 B 区域得到纠正，因此，它和异源的另一股 b

区域配对。

与两个替代基因之间简单的表达开关系统相比，这个系统所产生的变异实际上更广泛，因为基因转换可能不是涉及一个整体的基因，而只是其中的一部分，新表达的 *pil* 基因将包括两个不同拷贝的序列。据计算，这能使菌毛的抗原种类上升至 10^7 个品种。此外，由于淋球菌的自然转化（见第 6 章），包括 *pil* 基因的 DNA 在细胞间的转移，增加了其在逃避免疫反应时所具有的多功能性。

7.4.4 滑链错配导致的相位变化

一系列单核苷酸，如 AAAAAAAAAAAAA（同多聚片段），或者大于一个核苷酸的重复单元，称为多聚重复（如 **AT**ATATATATATAT 或 **GCC**GCCCCGCCGCC）。多聚重复提供了相位变化的另一种机制，在 DNA 复制时错配导致缺失或获得一种或更多的重复单位，这一过程叫做滑链错配（图 7.19）。当染色体复制时，在许多情况下，其序列不变，出现少数有不同数量重复单元的细胞，这可能会在翻译或转录水平影响基因的表达。如果重复在编码序列内，基因的阅读框将改变，导致翻译的过早终止，这种相位变化的机制，控制许多与毒力相关基因，包括奈瑟球菌（*Neisseria meningitidis*）产生的荚膜多糖，

相位开放

短序列重复(CTGG)

1 2 3 4

ATG CTGG CTGG CTGG CTGG TGACGGAT-------

M L A G W L V T D --------

在框中的基因：完全翻译

复制

CTGG

ATGCTGG CTGGCTGGTGACGGAT-----------

一个重复的缺失

ATGCTGG-CTGGCTGGTGACGGAT -----------

相位关闭

终止密码子

1 2 3

ATG CTGG CTGG CTGG TGACGGAT-----------

M L A G W *

丢失的重复单位已经在阅读框中改变

翻译提前终止

图 7.19 由滑链错配引起的相位变异

这些基因含有 4 个 4 碱基重复(CTGG)。在复制期间，假如这个重复的一个拷贝丢失(或者是添加)，阅读框将会改变，通常是导致蛋白质合成的提前终止

以及流感嗜血杆菌(*Haemophilus influenzae*),空肠弯曲菌(*Campylobacter jejuni*)和幽门螺杆菌(*Helicobacter pylori*)的脂多糖抗原的表达。

另一种重复序列可能位于调节区域内,并且基因长度的改变可能改变DNA和调节性蛋白或RNA聚合酶之间的相互作用。脑膜炎球菌外膜蛋白经历了这种形式的调节。在这种情况下,相位的改变引起启动子附近的Opc胞嘧啶碱基重复,并且会影响RNA聚合酶转录起始的效率,当这个重复的碱基数是10时,该基因不转录。在细胞这一区域含有14个胞嘧啶时,基因表达很弱,但是当有12个胞嘧啶时,是RNA聚合酶识别的完全表达的最佳间距。应该指出的是,这种相位的变化是很不寻常的,因为它不仅具有启动或关闭转录的调节功能,而且也控制蛋白质的产量。

7.4.5 不同的DNA甲基化介导的相位变化

上面所描述的所有的相位变化机制均来自于细菌基因组的变化,一种不同于这些相位变化的类型是表型变化,而基因组序列不变,换句话说,它是表观的。该机制的一个例子就是表达调控作用的*agn43*基因,该基因编码大肠杆菌主要相位变化的外膜蛋白(抗原43),虽然这个蛋白质的确切功能仍未知。根据其自身关联的特点,它介导大肠杆菌细胞的自身聚合和絮凝。

在抗原43相位变化中起关键作用的酶是脱氧甲基酶(DAM),它在DNA的N-6的位置添加了一个有甲基的GATC序列。DNA甲基化与错配修复(第1章)及修饰限制系统有关(第4章)。但是,DNA甲基化也能够改变与DNA之间调控蛋白的相互作用,因而被细菌用于以相变方式调节其中某些基因的表达。在通常情况下,与氧化应激有关基因的激活因子的调节子OxyR能够结合*agn43*的上游DNA区域发挥作用,在这种情况下显著地阻止了转录。(图7.20)。这种OxyR的结合位点包括3个GATC位点,这3个GATC位点是DAM甲基化的靶点。如果这3个位点被甲基化,OxyR不再能与DNA结合在一起,并且*agn43*基因表达。相反,在没有甲基化的情况下,OxyR能够与DNA结合并阻止转录。DNA复制后,这些位点随机甲基化,这导致相位变异,如一些细胞将表达抗原43,而另一些不表达。

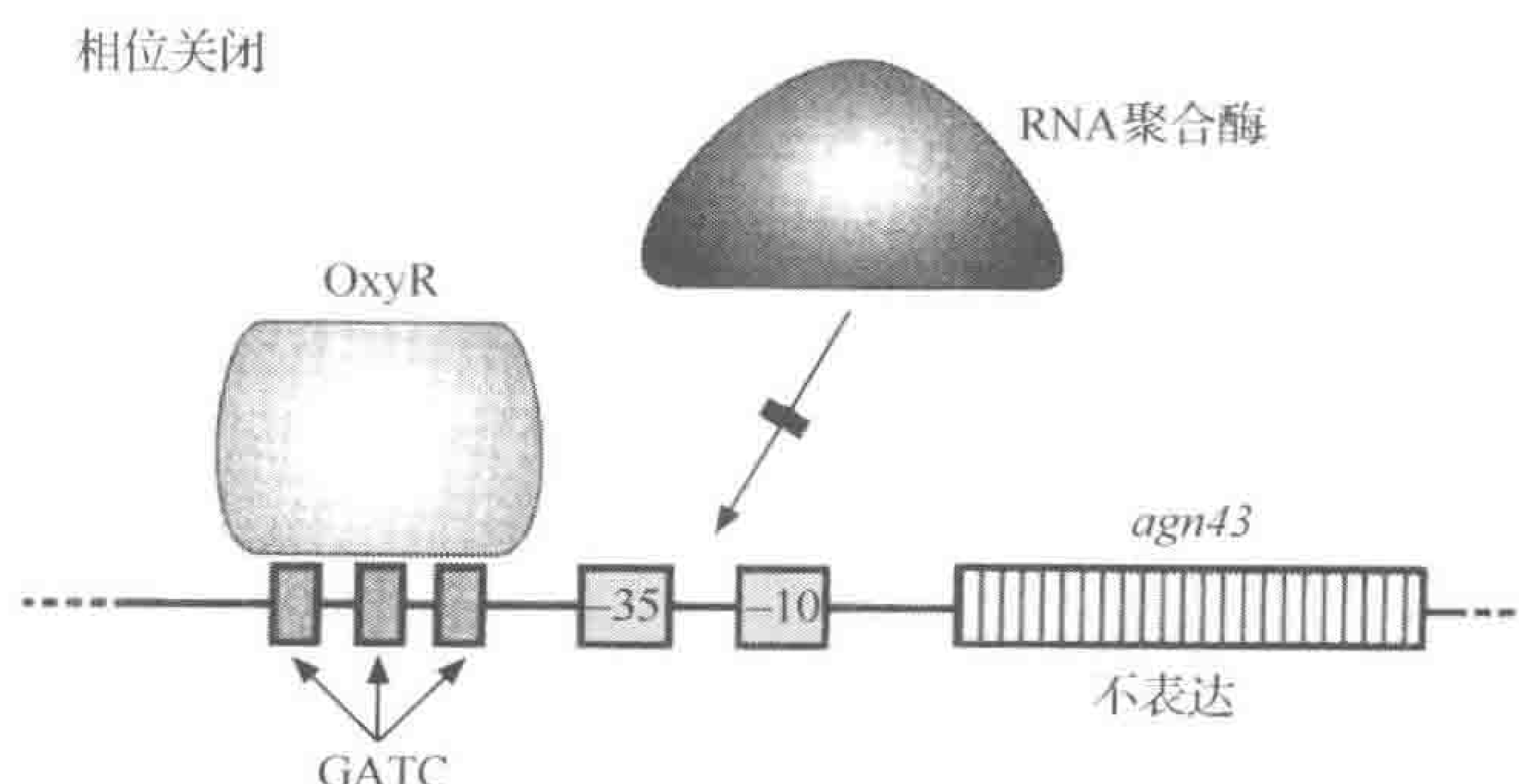


图7.20 由不同DNA甲基化导致的抗原43相位变异

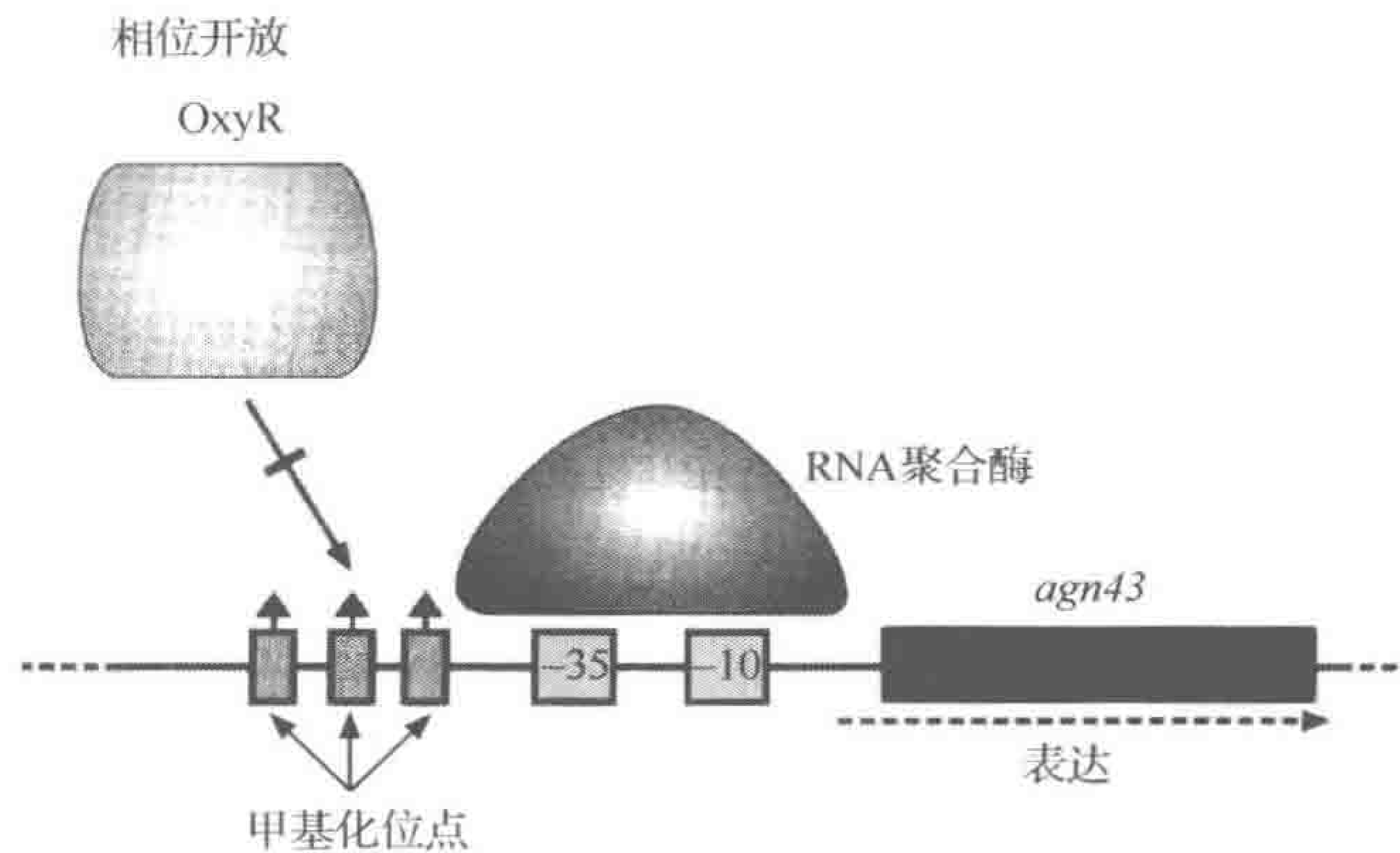


图 7.20 由不同 DNA 甲基化导致的抗原 43 相位变异(续图)

OxyR 可以阻止 *agn43* 基因转录。然而，如果在 OxyR 结合位点的 GATC 序列甲基化，OxyR 就不能结合，并且 *agn43* 基因将表达

7.5 规律性成簇的间隔短回文重复

和上述一些相关但具有不同效应的另一种类型的变化是由简称为 CRISPR 的规律性成簇的间隔短回文重复，其典型特征是具有一系列 30～40bp 序列的重复，用相似大小的独特的空间隔区序列隔开(图 7.21)，和这些相关的一些基因称为规律性成簇的间隔短回文重复相关(CAS)基因，被前导序列从重复中隔开。规律性成簇的间隔短回文重复序列在 40％的基因组序列中得到鉴定，在古细菌基因组序列中有 90％之多。

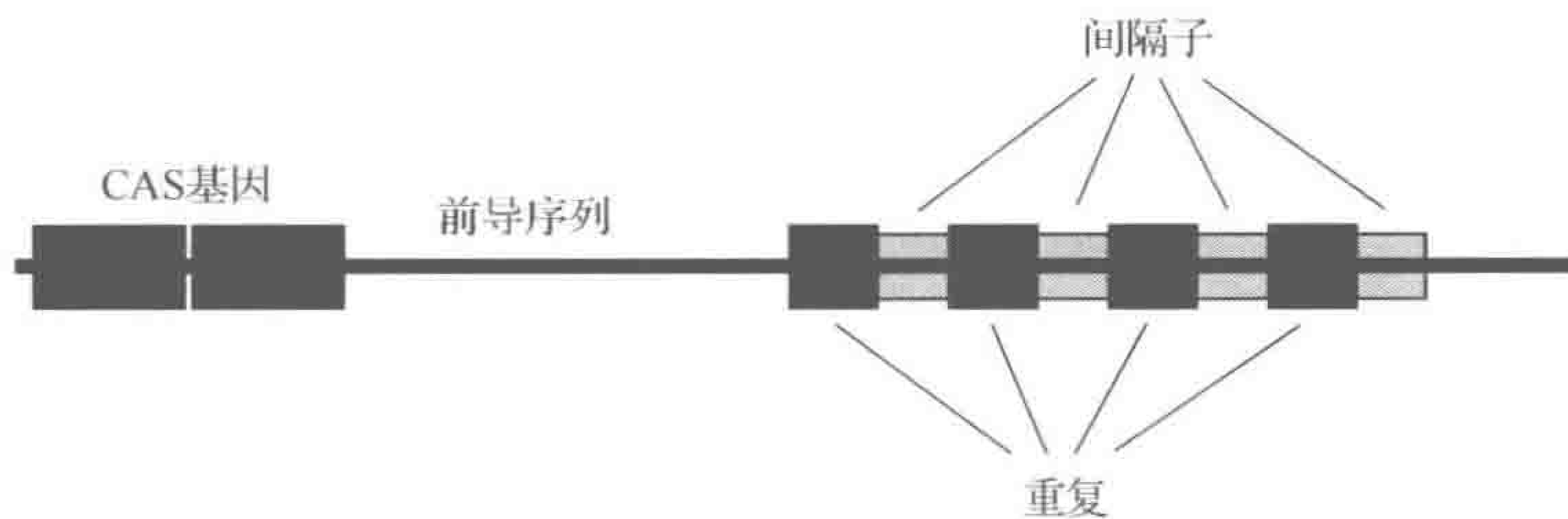


图 7.21 CRISPR 区域结构

CRISPR 含有重复序列，这些重复序列已被识别或者将被识别，并用间隔序列分类，这些间隔序列是独一无二的

在同种细菌不同菌株之间，重复间隔区单位的数目和间隔区的同一性差异很大，这样的特性被用于细菌的定型系统中，特别是结核分枝杆菌(*Mycobacterium tuberculosis*)的分型中。然而，这个区域多年来仍然保持着神秘性，如果有的话，包括间隔区的起源和它们的目的都是未知的。已证实许多间隔区序列来源于噬菌体和质粒，而且 CRISPR 区域在细菌防御中发挥着抵抗噬菌体攻击的作用。在噬菌体感染之后，在存活的具有噬菌体抗性的嗜热链球菌(*Streptococcus thermophilus*)的突变体中发现了额外的重复间隔区，该间隔区处于区域的前导区末端。该间隔区域的序列与感染性的噬菌体相匹配。将对噬菌体敏感的菌株加上这些重复间隔区后，这些敏感的菌株变成了对噬菌体具有抗性的

菌株。

尽管作者清楚 *CAS* 基因参与了这个过程，但这个间隔区是如何被加上去的，以及细菌如何变得具有抗性，这些仍然没有被强有力的证据予以证实。噬菌体抗性被想象成包括 CRISPR 区域的转录，接下来转录本被酶切割成每一个都包括一个间隔区的多个小片段。如果细胞被噬菌体感染，而噬菌体的染色体内拥有与一个特别间隔区相对应的序列，然后间隔区与噬菌体的 mRNA 相结合，导致噬菌体的 mRNA 减少。这样的模型是与 RNA 沉默或 RNA 干扰相关的，真核细胞抵抗 RNA 病毒的感染和广泛的基因调控现象也包括这样的机制。尽管关于这些机制的研究主要存在于真核细胞中，现在在细菌中关于 RNA 调控机制也被认为是非常重要。在第 3 章中已经对这样一些例子进行了讨论。

8 遗传修饰：细菌潜能的开发

基因修饰已经成为与体外基因操作技术(基因克隆)相关的术语，但是，出于各种目的，在应用该术语很久以前就已经使用基因修饰或操作的方法来改造细菌了。体外基因操作技术与较老的、体内的方法相比，主要的不同是体内方法对被选择的、拥有合乎需要特征(也就是表型)的突变体和变异体的遗传本质知之甚少，甚至毫无所知。而体外遗传操作却是相反的过程，通过改变特定的基因(或插入外源基因)，希望能够出现有用的表型。对各种基因的功能和调节的认识以及它们与整个细胞基因组成的关系的认识程度将影响着取得成功的机会。尽管目前研究的重点主要集中在体外实验方法上，但有时，特别是研究对象遗传背景不是很清楚时，体内实验方法实际上更简单、快捷。

在本章中，将回顾可以用来产生有用菌株的可行方法，有用的菌株包括增加天然产物的形成如抗生素，非细菌产物产量如人生长激素，或者是抵抗感染性疾病的疫苗开发。在第 9 章中，将关注遗传学在加深对细胞生理和生化过程认识中的作用，而在第 10 章中，将继续关注基因组测序和生物信息学在加深对细胞生理和生化过程认识中的作用。在这 3 章中，很多概念和技术有一定联系，许多材料也被一定程度的反复引用。

8.1 菌株的改良

自然进化主要基于下列 3 个过程：产生变异菌株，根据需要的特征筛选变异菌株及通过菌株间的遗传物质交换使某些特征再分布。通过菌株改良的方法得到有商业价值的微生物，通常是增加产物的合成，主要包括在这一进化过程中的前两个环节。

8.1.1 变异的产生

在前面章节中，了解到变异可自然发生，但频率较低。为了实际应用，需要通过使用诱变剂的方式增加变异的频率。早期研究主要使用如 X 线和氮芥子这样的强诱变剂，但是极有可能产生一些额外的不期望的效果，同时其使用也缺乏安全性。随后，较温和而容易控制的诱变剂如紫外线辐射逐渐被应用(见第 2 章)。

对受影响的基因来讲，基因突变极可能是有害的，所以可能期待分离出产量增加的菌株比较困难，但这只是部分事实。之后发现，可以通过清除目标产物产生的调节途径或是降低其他减少目标产物积累的代谢活动来增加产物合成。在对产生途径和调控一无所知的情况下，可以根据经验对所需表型进行筛选，获得所需要的突变体。

8.1.2 目标变异菌株的筛选

从进化的观点看，因为改良菌株的适应性更佳，而被选择出来。同样，在微生物遗传学中，所谓的筛选意味着使环境只适合所需菌株生长，对抗生素抗性突变体的筛选就是一个最清楚的例子。但是，通过菌株改进来获得有商业用途的微生物时，这样做几乎是不可能的。产生更高水平的抗生素并不能为生产菌株提供可筛选的优势，而是需要分离出每一个单克隆，并且测试每个克隆的产物水平。这样一种过程最好的定义为"筛选"，但是非常容易混淆的是这一过程通常被称为"选择"。

对于商业化的菌株来说，除了产物的产量外，其他一些特性也很重要，如生长速度，底物的利用，对不同发酵条件的反应以及不产生不期望得到的副产物。这些副产物不仅会降低所需材料的产生，还有可能污染最终的产品，这样就增加了下游环节的成本（即产品的分离和纯化）。

8.2 初级代谢产物的过量产生

在微生物生物技术中，代谢物通常分为初级代谢物和次级代谢物。初级代谢物，是指那些作为细胞正常生长的一部分，在新陈代谢过程中有一定作用的一类物质（如氨基酸）。而次级代谢物（如抗生素），一般产生于细胞稳定期，并且对细胞的生长似乎并不起核心作用。

8.2.1 简单途径

初级代谢物可以看做前体底物（S）经过一系列反应生成的最终产物（P）（图 8.1）。可能限制 P 产量的第一个主要因素是前体底物（S）的供给量，提高细胞内 S 的水平将会增加该通路终产物的产量。但是，由于 S 同时被细胞内其他代谢途径利用，改变 S 在细胞内的可利用水平是很难的。

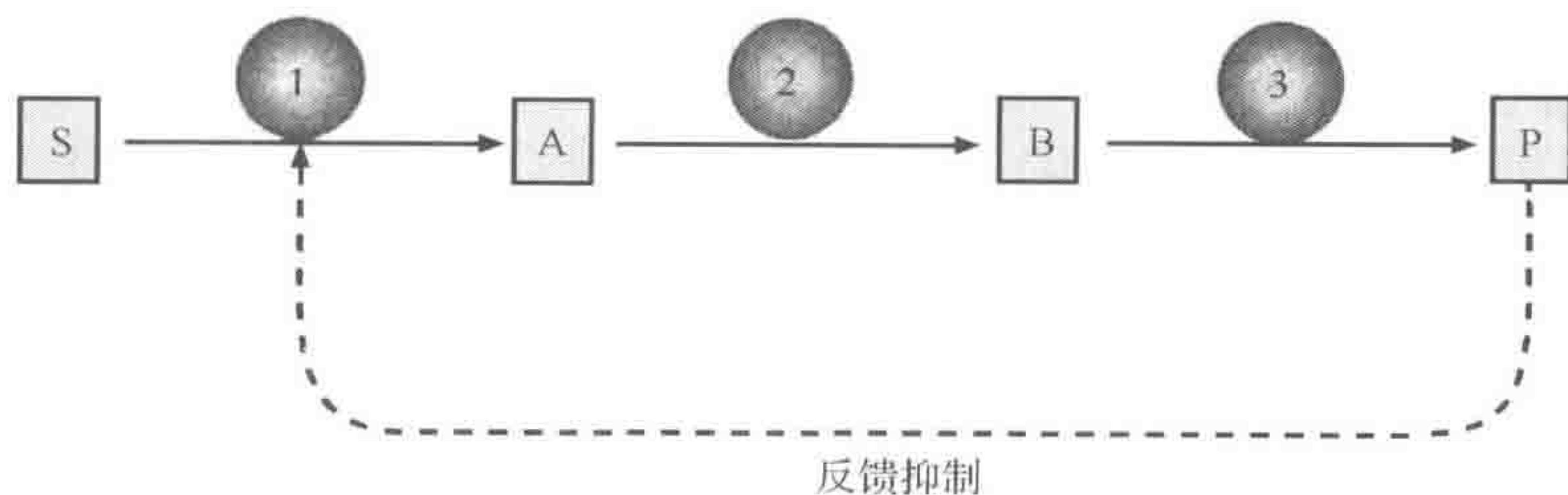


图 8.1 初级代谢产物的调节：一种简单无分支的通路

其次，讨论各个酶促反应的速率。在通常情况下，酶促反应的速率受酶分子数目、酶的催化活性、酶与底物的亲和力及酶底物浓度的影响。理论上，可以通过改变酶的结构来增加酶的最大活性和（或）它与底物的亲和性，但实际上，这样的突变体却非常稀少。更有效的办法是通过改变酶的启动子来增强该基因的转录，从而提高酶的产生速率。然而，假

设整个系统的产量仅取决于酶的数量，这就过分简单化了。如果增加图 8.1 中酶 2 的数量，一方面将可能会提高系统的产量，但另一方面，中间产物 A 的量将会成为限制（因素），或者酶 3 将不能有效地催化增加的中间产物 B。

反馈调节

第三个（通常是最重要的）因素是通路的调节。一条通路可能是高产出的，但随着产物数量的增加，反馈作用将会抑制产物的生成，这可能是由阻遏或抑制导致的。阻遏是酶的生成减少造成的，而抑制是由酶活力的降低形成的。其中之一或二者的作用意味着试图增加 P 的产量将直接导致通路的关闭。

因此，如果筛选过量表达终产物的克隆，最有可能获得在反馈调节通路中有缺陷的突变体。如果以色氨酸的生成来作为一个例子，（在大肠杆菌中）5 个基因的整个操纵子在色氨酸存在的情况下都被抑制（第 3 章）。破坏这些控制的突变会解除细菌对色氨酸的反应，导致产物合成量的增加。

另外，反馈抑制（通常）是由于终产物结合到酶的某一位点引起酶蛋白构象改变而不能正常发挥其催化作用。尽管有可能获得抗反馈抑制的突变体，但与使抑制蛋白失活相比，对酶进行改造要精细得多，所以，想通过随机筛选获得突变体非常困难。

抗代谢物

有时，可以通过添加抗代谢物来分离抗反馈抑制的突变体。有许多代谢途径终产物的类似物对机体有致死效应，这是由于该类似物在反馈作用中替代了真正的产物，但却不能替代真正产物的代谢作用。例如，色氨酸类似物 5-甲基色氨酸可以抑制色氨酸通路（从而没有色氨酸生成），但却不能够在蛋白质合成中替代色氨酸。因为 *trp* 操纵子可以表达，色氨酸将被生成并被用于蛋白质合成，缺乏反馈调节的突变体可以在 5-甲基色氨酸存在的条件下生长。因此，通过将诱变的培养物接种到含有抗代谢物培养基中可以筛选调节型突变体。

8.2.2 分支途径

许多氨基酸是分支途径合成的产物，而非简单途径合成的。在图 8.2 中，副产物 R 的产生分流了生成必需产物 P 的资源。缺乏酶 4 的突变体，不能够合成 R，将有可能生成更多的产物 P。缺乏 R 的另一个好处是分支途径中通常存在协同（或多价）抑制，在上述例子中，只有在当 P 和 R 总量达到一定时，酶 1 的活性才会被抑制。如 *ilv* 操纵子只有在亮氨酸和缬氨酸同时存在时才被抑制（二者都是分支途径的终产物），而缬氨酸与异亮氨酸（的生成）需要同一个酶。另一个例子是将在后面详细讨论的赖氨酸的生成。

总体来看，在图 8.2 中，如果非目标的副产物 R 对细胞生长为所必需的，那么缺失酶 4 的突变体将是 R 营养缺陷型的，并且可以通过复制平板而被筛选出来。突变体的培养需要在生长培养基中加入 R。这样也许会反馈抑制阻遏酶 1。然而，添加有限量的底物通常是可能的，这样可以保证细菌生长却不引起反馈作用。

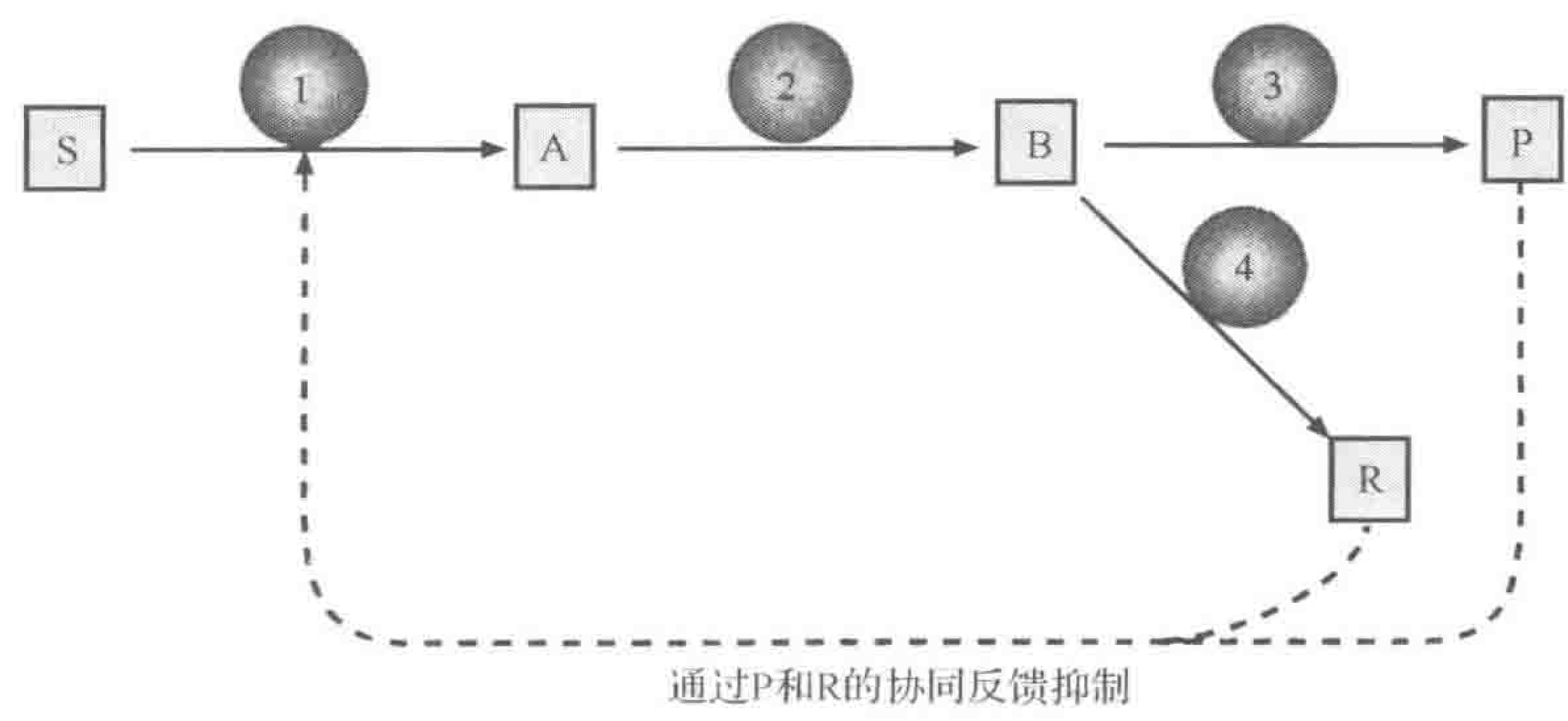


图 8.2　初级代谢产物的调节：一种有分支的通路

赖氨酸合成

赖氨酸广泛用于谷物类动物饲料的添加，其商业化生产是一个特定的例子。图 8.3 是谷氨酸棒杆菌（*Corynebacterium glutamicum*）中赖氨酸合成途径的简图，与其他细菌一样，赖氨酸合成的起始阶段与苏氨酸、异亮氨酸和甲硫氨酸的合成是相同的。

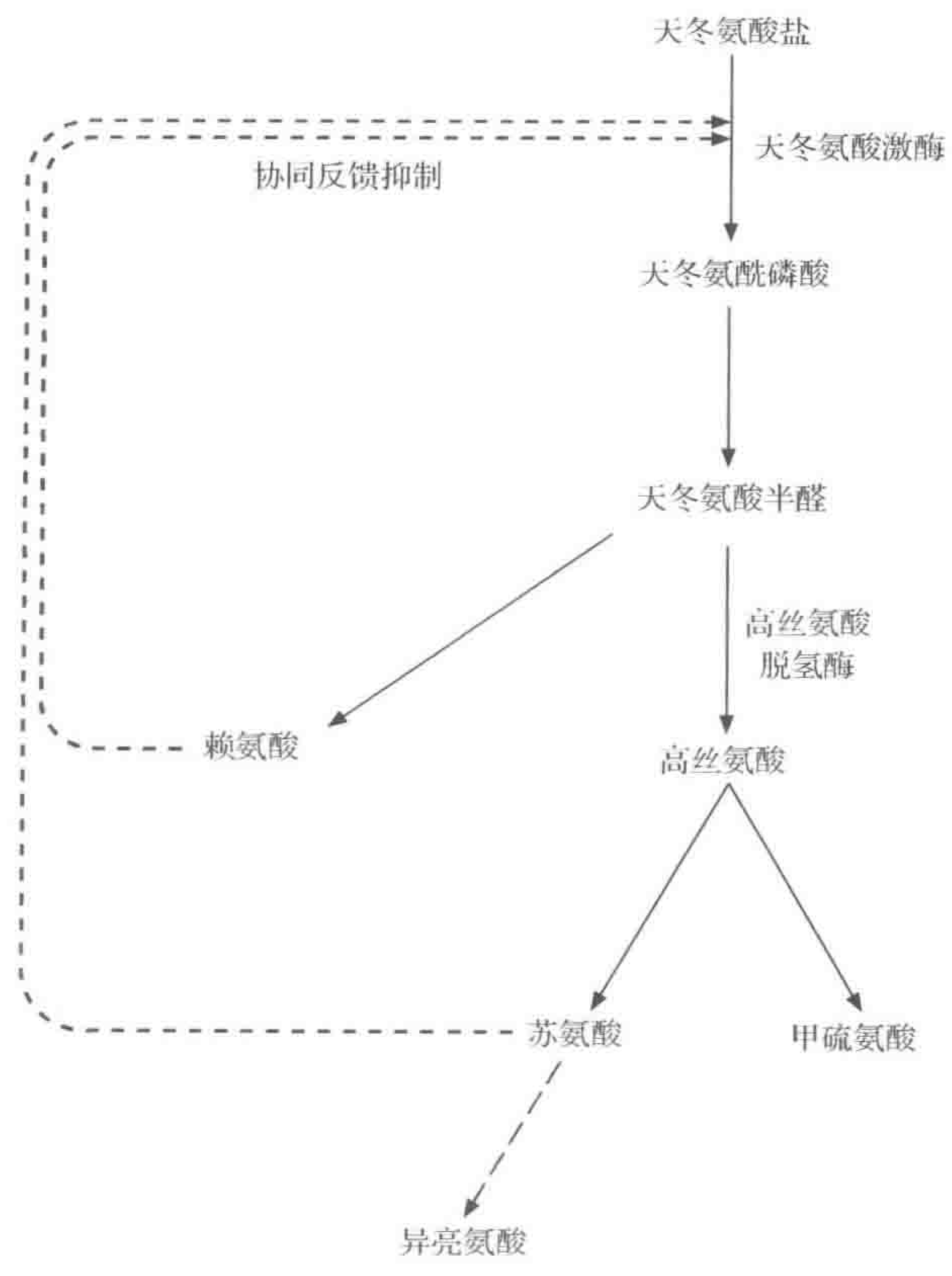


图 8.3　谷氨酸棒杆菌中赖氨酸的合成

缺乏高丝氨酸脱氢酶的谷氨酸棒杆菌突变体是营养缺陷型的，但可以在添加有高丝

氨酸(或苏氨酸和甲硫氨酸的混合物)的培养基中生长。这样的突变体可以利用部分与其他氨基酸代谢不同的途径合成大量的赖氨酸(超过 50g/L),而且不受反馈调节。共同途径中的第一个酶(天冬氨酸激酶)受赖氨酸和苏氨酸的协同反馈调控。但在营养缺陷型突变体中,在培养基中加入有限的高丝氨酸,苏氨酸浓度太低,以至于不会引起对天冬氨酸激酶的抑制。

8.3　次级代谢产物的过量产生

抗生素是微生物产生的最重要的一种次级代谢物。第一个真正(意义上)的抗生素——青霉素,是真菌(青霉菌)的产物,但天然抗生素的主要来源是丝状放线菌[特别是链霉菌属(*Streptomyces*)]。通常情况下,自然菌株产生的抗生素水平都很低,不能实现商业加工,所以,需要进行菌株的改良。然而,合成次级代谢产物的代谢途径比初级代谢物更加复杂多样,因此,当对代谢途径或调节过程的认识还不深入时,菌株的改良过程只能靠经验进行。

这将是一个耗时的过程,但并非无法克服。通过对大量克隆进行筛选,可以获得表达产物量提高的变异克隆,之后再进行新一轮的诱变和筛选。重复循环这一过程,可以使菌株(产生)抗生素水平比原始菌株的产量高。

值得强调的是,大多数用于生产抗生素的菌株很大程度上或是完全来自这样的凭经验的筛选过程。尽管现在许多代谢途径已被阐述清楚,基因组测序加深了对相关过程的理解,随机突变和筛选仍将可能是获得高水平产物克隆的主要方法。另外,对各代谢途径及其调控的详细了解,有助于利用各代谢途径产生各种需要的抗生素衍生物。在这样的条件下,从系统生物学和代谢组学(第 10 章)获得的知识,将有助于从整个细胞水平加深对细菌新陈代谢的学习,并更进一步地认识代谢途径的调节。

8.4　基因克隆

经典的遗传技术(体内)主要受两个因素所限制。首先,这些技术只能用于生物体已有的遗传组分,也就是它们仅限于用于天然存在的基因或经少量修饰的天然基因。运用这些技术在生物体中产生一个外源的产物是不可能的。其次,这些经典的技术只能针对表型来开展工作,也就是说筛选突变体只能根据观察到的生物体的表型特征来筛选。这些不足严重限制了可被筛选的突变种类。

基因克隆(也叫做体外遗传操作或遗传工程)的出现在上述两方面都引起了巨大变化。这些技术的基础是利用酶进行 DNA 片段的剪切和重组。这样,外源 DNA 片段可以被插入到载体(质粒或噬菌体)中,从而在细菌中复制。

两者更进一步的区别在于,在基因克隆中,遗传变异的性质通常是已知的,尽管表型或许不可预测,而在体内方法中,可以知道表型但要清楚确定起作用的遗传突变的本质却很困难。基因组测序的最新进展为这一问题的解决提供了一条可能的途径。只需要简单地挑出感兴趣的表型的突变体,然后再进行整个基因组的测序,就可以精确地发现是哪个基因发生了突变。

8.4.1 DNA 的剪切与连接

限制性内切核酸酶具有对 DNA 片段特异的位点的切割能力(框 8.1),是基因克隆中的一个关键因素,因为它可以获得特定的 DNA 片段。而获得的片段又可以被 DNA 连接酶连接起来。由同一个限制酶产生的任意两个末端都可以通过这种方式连接起来,但不能够将一个 *Eco*RⅠ产生的片段与 *Bam*HⅠ产生的片段连接(然而,可以将一个 *Bgl*Ⅱ产生的片段与 *Bam*HⅠ或 *Sau*3A 产生的片段相连接,虽然这些酶识别不同的位点,但它们可以产生相同的黏端)。同时,DNA 连接酶还可以连接平端。尽管连接效率较连接黏端低,但由于不要求连接片段都是由同一个酶酶切产生,平端连接也很实用。

框 8.1 限制性内切核酸酶

限制性内切核酸酶(限制酶)可以在特定的位点切割 DNA。在表中列举了一些例子。酶的名称,是由产生这种酶的生物的属名的第一个字母和种名的前两个字母,再加上其他字母或数字来组成(因为一个物种可能产生很多种不同的限制酶)。

在 DNA 中,如果 4 种碱基出现概率均等(也就是 G+C 含量占 50%,在大肠杆菌中比较接近这个比例,而在其他生物中则通常差异很大)且随机分布,那么,一个特定的 6 个碱基的序列出现的平均概率就是每 4^6(4096 碱基)个碱基出现一次。因此,可以预期像 *Eco*RⅠ可以把 DNA 切割成平均大小约 4kb 的片段。而识别位点为 4 个碱基的酶,预期平均片段长度为 4^4(256)个碱基。

限制酶的识别位点通常为回文序列。如果是将双链 DNA 序列旋转(保持链的方向与旋转前一致),那么得到的序列和旋转前是完全一样的。

*Eco*RⅠ在 G 与 A 碱基之间不对称地切割 DNA,如表中斜线(/)所示。这样,就在两个末端产生了 4 个不配对的碱基(本情况是 5′-AATT)。这些单链区域是互补的,可以与带有同样单链的其他片段(或者同一片段的另一末端)配对。这样的末端被称为黏端。

还有一些酶,如 *Pst*Ⅰ,在切割片段的 3′端产生 4 个碱基的黏端,而像 *Sma*Ⅰ,在识别序列中心进行切割,在被切割的片段两端产生平端。

下表列出了一些限制性内切核酸酶的例子

酶	识别位点	碱基数	末端类型	酶的来源
*Eco*RⅠ	G/AATTC	6	5′黏端	大肠杆菌(*E. coli*)
*Hind*Ⅲ	A/AGCTT	6	5′黏端	流感嗜血杆菌(*H. influenzae*)
*Bam*HⅠ	G/GATCC	6	5′黏端	解淀粉芽孢杆菌(*Bacillus amyloliquefaciens*)
*Pst*Ⅰ	CTGCA/G	6	3′黏端	斯氏普罗威登斯菌(*Providencia stuartii*)
*Sma*Ⅰ	CCC/GGG	6	平端	黏质沙雷氏菌(*Serratia marcescens*)
*Sau*3A	/GATC	4	5′黏端	金黄色葡萄球菌(*Staphylococcus aureus*)
*Alu*Ⅰ	AG/CT	4	平端	酵母菌用藤黄节杆菌(*Arthrobacter luteus*)

如果将片段连入一个质粒或噬菌体这样具有复制能力的载体，就可以通过将连接后的载体转化进入合适的宿主，实现克隆片段的扩增。使用质粒载体，在琼脂板上挑出单克隆，用这个克隆扩增出大量细菌培养物，在每一个细胞中都有原 DNA 片段的一个拷贝。严格地讲，克隆是指从众多转化细胞的混合物中挑选出单克隆（或从噬菌体侵染物中挑选到的单个噬菌斑）的过程。然而，基因克隆通常指整个过程，或仅仅是指 DNA 片段与载体重组的过程。与其他技术相比，这项技术最大的优势在于，首先，它对 DNA 片段的来源没有限制；DNA 只是 DNA，可以被克隆并在大肠杆菌中扩增，不论它的来源以及序列如何。其次，不需要对克隆片段进行纯化。通过一定的方法可以知道哪个细菌克隆含有感兴趣的基因，非常简单地就可以分离纯化它们。

8.4.2　质粒载体

基因克隆使用的质粒载体的许多特点，可以通过一个在基因克隆中广泛使用的载体 pUC18（图 8.4）进行说明。

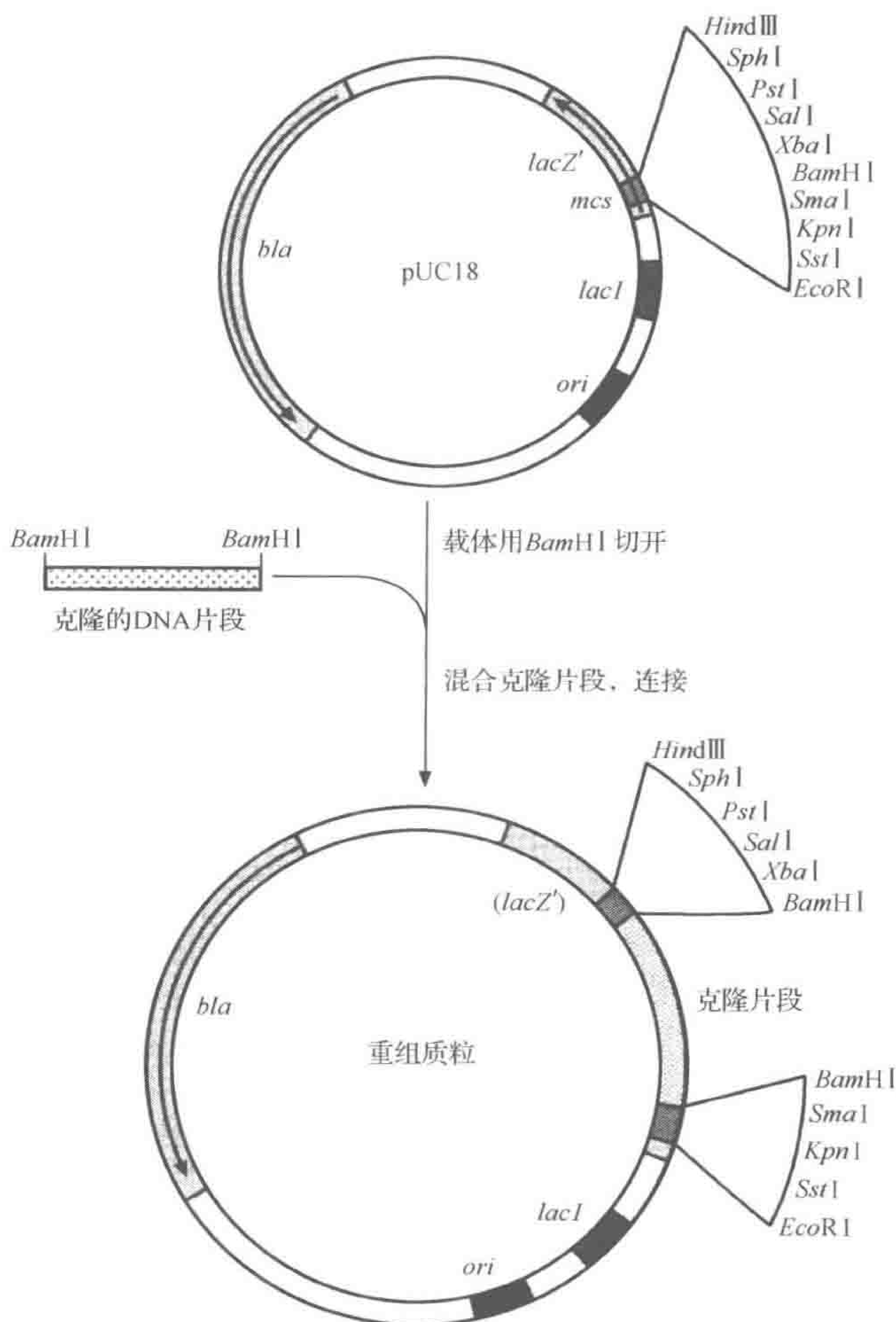


图 8.4　质粒克隆载体 pUC18 的结构和用途

复制原点

首先，质粒要能够在选定的宿主（本情况下是大肠杆菌）中扩增。用于构建 pUC18 质粒的复制子与 ColE1 质粒有亲缘关系，这是一段包含复制原点（*ori*）的区域。因此，质粒就可以按照 ColE1 质粒的方式进行复制（第 5 章）。

筛选标志

其次，需要能够对获得外源质粒的细胞（转化细胞）进行筛选。因此，在克隆质粒载体构建时，通常要带有一个或多个抗生素抗性基因；pUC18 带有一个 β-内酰胺酶基因，可以对氨苄青霉素产生抗性。

同时，pUC18 上还带有另一个标志，β-半乳糖苷酶基因（严格地讲，pUC18 只编码 β-半乳糖苷酶的一个片段，剩下的部分由宿主基因编码，从而形成有活性的酶，也就是 pUC18 补充了无活性的宿主基因）。将 pUC18 转化进合适的宿主可以产生有活性的 β-半乳糖苷酶，转化克隆可以通过 β-半乳糖苷酶水解发色底物 X-gal 产生蓝色产物而被检测。因此，有 pUC18 的克隆是蓝色，而没有 pUC18 的克隆则是白色。然而，这并不是选择标志，因为两种类型的细胞都可以在选择性培养基上生长，但是，这样提供了一种额外而有用的区分两种细胞的方法（见后文插入失活）。

克隆位点

对于克隆载体来讲，第三个必须具有的特征就是有合适的识别位点，可以被一个或多个限制性内切核酸酶切割。这样，环状的质粒可以在这个位置被打开，（被限制酶切开的）末端可以与克隆 DNA 片段末端连接起来（图 8.4）。

例如，一个将被克隆的 DNA 片段是由 *Bam*H Ⅰ 酶切产生，把它与由同一个酶线性化的 pUC18 混合，DNA 连接酶将把这些片段中由 *Bam*H Ⅰ 产生的末端连接起来，这样就形成了图 8.4 中所示的重组质粒。

pUC18 载体中有一段合成的 DNA，包含许多限制酶的识别位点。这样的多克隆位点（mutiple cloning sit，MCS）为限制酶的选择提供了很大灵活性。同时，质粒经过改造，使这些位点在质粒上是唯一的，也就是除多克隆位点上的识别位点外，载体上没有被这些酶识别的位点。

插入失活

载体上克隆位点的位置很重要。DNA 片段的插入通常会导致插入点基因的失活。因此，克隆位点不能在质粒的必需区域。在图 8.4 中可以发现，多克隆位点在 β-半乳糖苷酶基因内的起始部位（但不会使基因失活）。如果一个 DNA 片段插入到多克隆位点，通常会通过干扰翻译或改变阅读框而抑制 β-半乳糖苷酶的产生；这就被定义为插入失活。真正的重组子克隆将会在含有 X-gal 的培养基上呈现白色，而与含有原始的 pUC18 克隆呈现的蓝色相区别。

转化

重组质粒构建好后，必须将其转化到宿主细胞中。在第 6 章中已经介绍的天然的转化系统，不适合质粒的转化。一方面是因为环状的质粒不能通过该系统进行很好的转化，另一方面就是在天然状态下，在包括大肠杆菌在内的许多细菌中，细胞可以摄取（外源）DNA 的感受态不能自然地发生。对于这样的细菌，转化要求人工诱导感受态。有许多方法可以达到这一目的，其中最简单的是用冰冷的氯化钙溶液反复洗细胞（图 8.5）。感受态细胞与 DNA 溶液混合后被施以热休克处理，如 42℃热激 1～2min，再放回冰上。加入生长培养基，在 37℃温育使新获得的 DNA 表达，最后涂布在合适的培养基上。

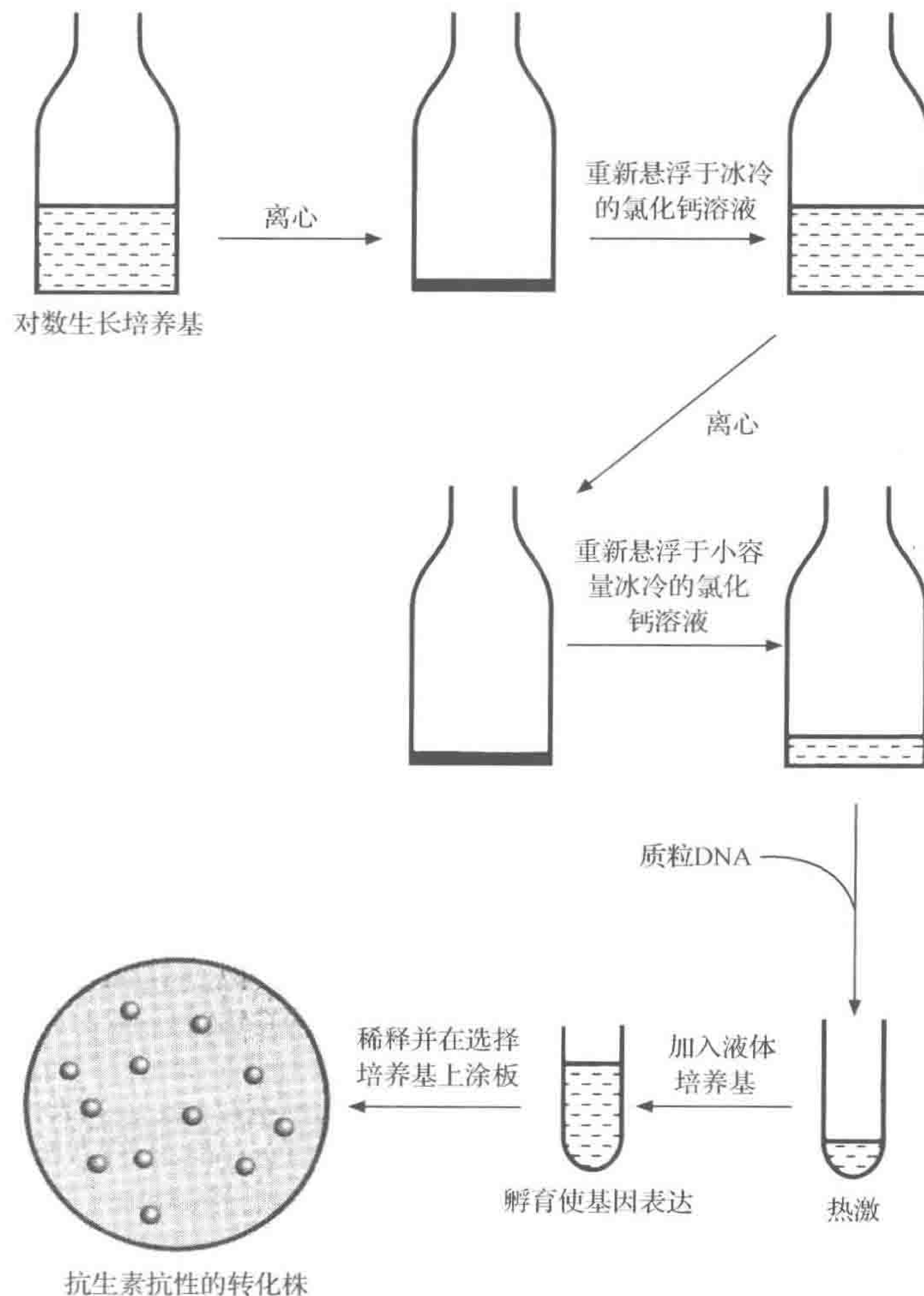


图 8.5　大肠杆菌质粒转化的基本程序

大量的工作致力于对大肠杆菌转化系统的优化，包括改进特殊菌株，使其具有很高的转化效率，每微克 DNA 可以形成 10^9 个转化体。即使在这样的系统中，也只有一部分宿主细胞可以摄取 DNA，因此，仍然需要有一个前文所述的好的筛选标志。

其他类型的转化，特别是电转化，对除大肠杆菌以外的其他宿主细胞非常有用，将会在本章后面部分进行介绍。

8.4.3 λ噬菌体载体

λ噬菌体来源的载体在大肠杆菌基因克隆中被广泛使用。尽管分子生物学的最新发展降低了其重要性，一个简单的介绍还是有用的，尤其是其中包括了许多λ噬菌体的生物学特性。

插入型载体

在简单的一类λ噬菌体载体即插入型载体中，只有在特定限制酶的单个酶切位点，克隆片段被插入该位点(图 8.6)。

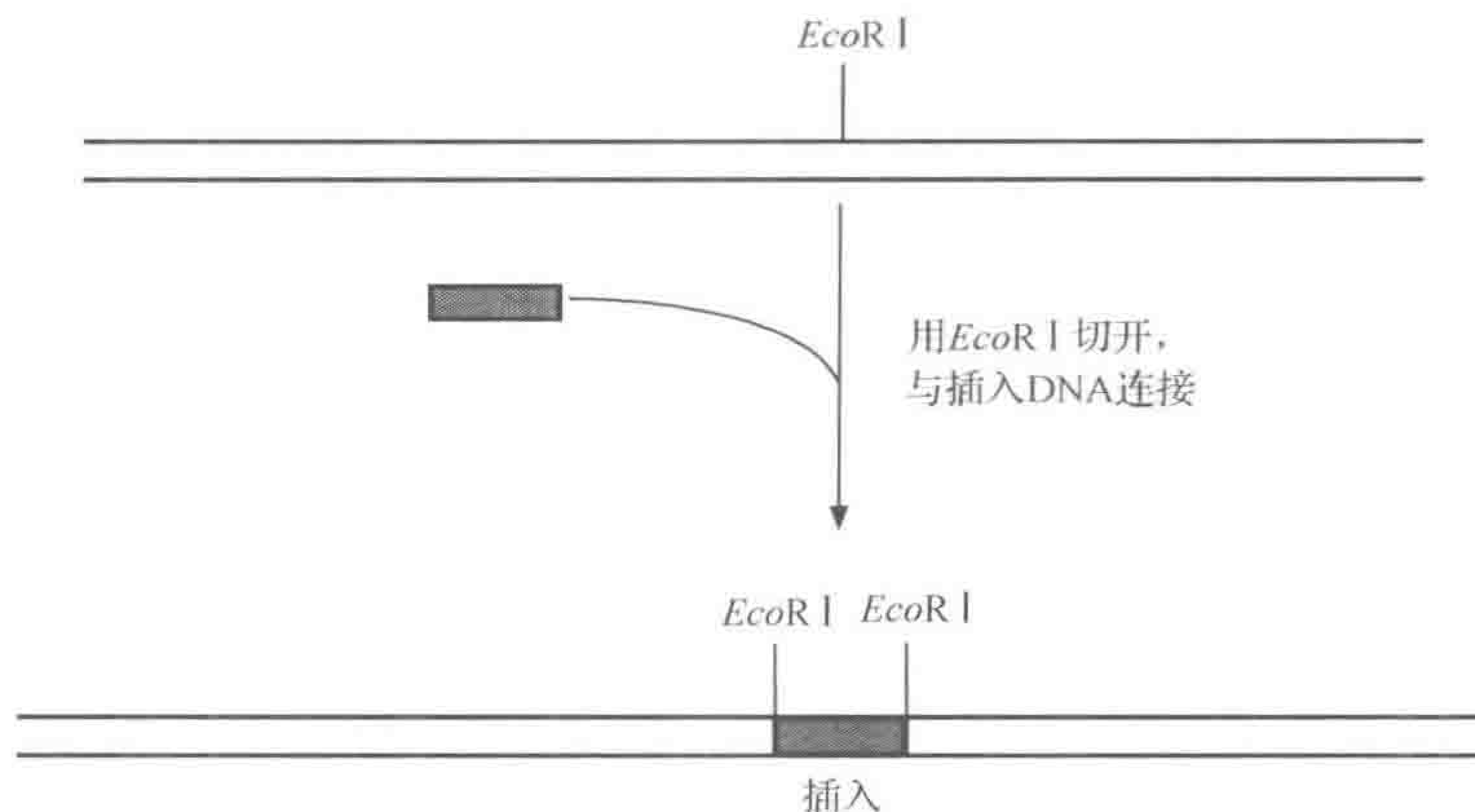


图 8.6 λ插入型载体的结构和用途

与质粒载体相比，噬菌体最大的不同在于，DNA 进入细菌细胞后，寻找的是菌苔中的噬菌斑(见第 4 章)，而不是选择性抗生素抗性克隆。另一个重要的不同是得到的吸收 DNA 细胞的方式。在λ噬菌体的自然溶菌周期中，从多倍单位长度的 DNA 分子剪切下来的单位长度 DNA 片段，被包装进如第 4 章所描述的空的噬菌体头部中，产生能够将目标 DNA 注射进宿主细胞的病毒颗粒。在体外，使用含有噬菌体头部和尾部以及所需酶的细胞提取物也可以实现包装，体外包装这一过程产生的病毒颗粒能够注入 DNA 进入宿主细胞。

使用插入型载体时，克隆片段大小受λ噬菌体自然包装的限制。λ噬菌体基因组中与溶源菌生长无关的区域可以被删除，这样可以为插入的片段提供空间。然而，删除的片段长度不能超过 10kb，或载体的长度不能小于最小包装长度，因为载体本身将不能存活。遗憾的是，λ噬菌体最主要用于构建基因组文库(见下文)，这通常要求插入片段尽可能的大。

置换型载体

λ噬菌体载体的克隆容量可以通过其他方法提高。在图 8.7 所示的例子中，载体上

有两个限制性位点而不是一个。酶切后，载体 DNA 生成 3 个片段：左臂、右臂和中间的填充片段。填充片段不含有噬菌体生长必需的基因，因此可以被分离并去除。左臂和右臂由于长度太小不能包装形成噬菌体颗粒，因此在这里需要有一段长度至少 7kb 而又不大于 22kb 的插入片段。因此，这样的载体不仅增加了噬菌体载体的克隆能力，而且可用于具有合理长度的较大片段的克隆。这种类型的载体被称为置换型载体。

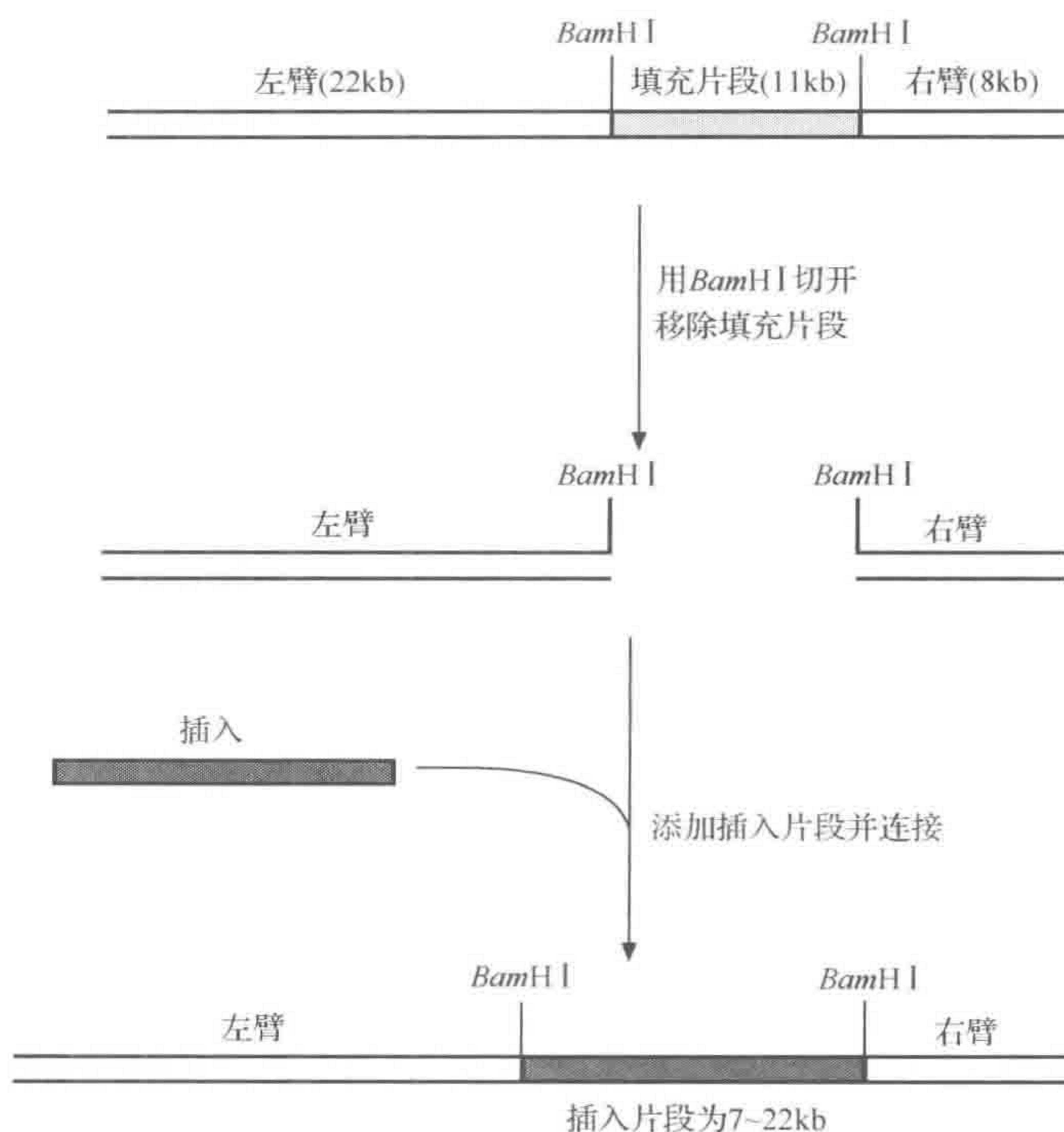


图 8.7 λ置换型载体的结构和用途

8.4.4 大片段的克隆

λ噬菌体 DNA 的包装只要求 *cos* 位点周围的区域(见第 4 章)，这样就可以对一类叫黏端质粒的载体进行体外包装(图 8.8)。这类质粒就是在一般的质粒中引入了 *cos* 位点。黏端质粒本身太小而不能被有效地包装。但是，如果插入片段很大(如 40kb)，重组的黏端质粒将大到可以包装产生存活病毒颗粒的程度。这些病毒颗粒可以感染敏感宿主，但在这种情况下，只能寻找具有抗生素抗性的克隆而不是噬菌斑。

更大片段的克隆，可以使用噬菌体 P1 载体，这种载体能容纳外源片段的长度为 100kb。也可用被称为细菌人工染色体(BAC)系统。这一系统是在 F 质粒基础上构建的，可容纳大于 300kb 的插入片段。这样的载体在基因组测序中特别有用(第 10 章)。

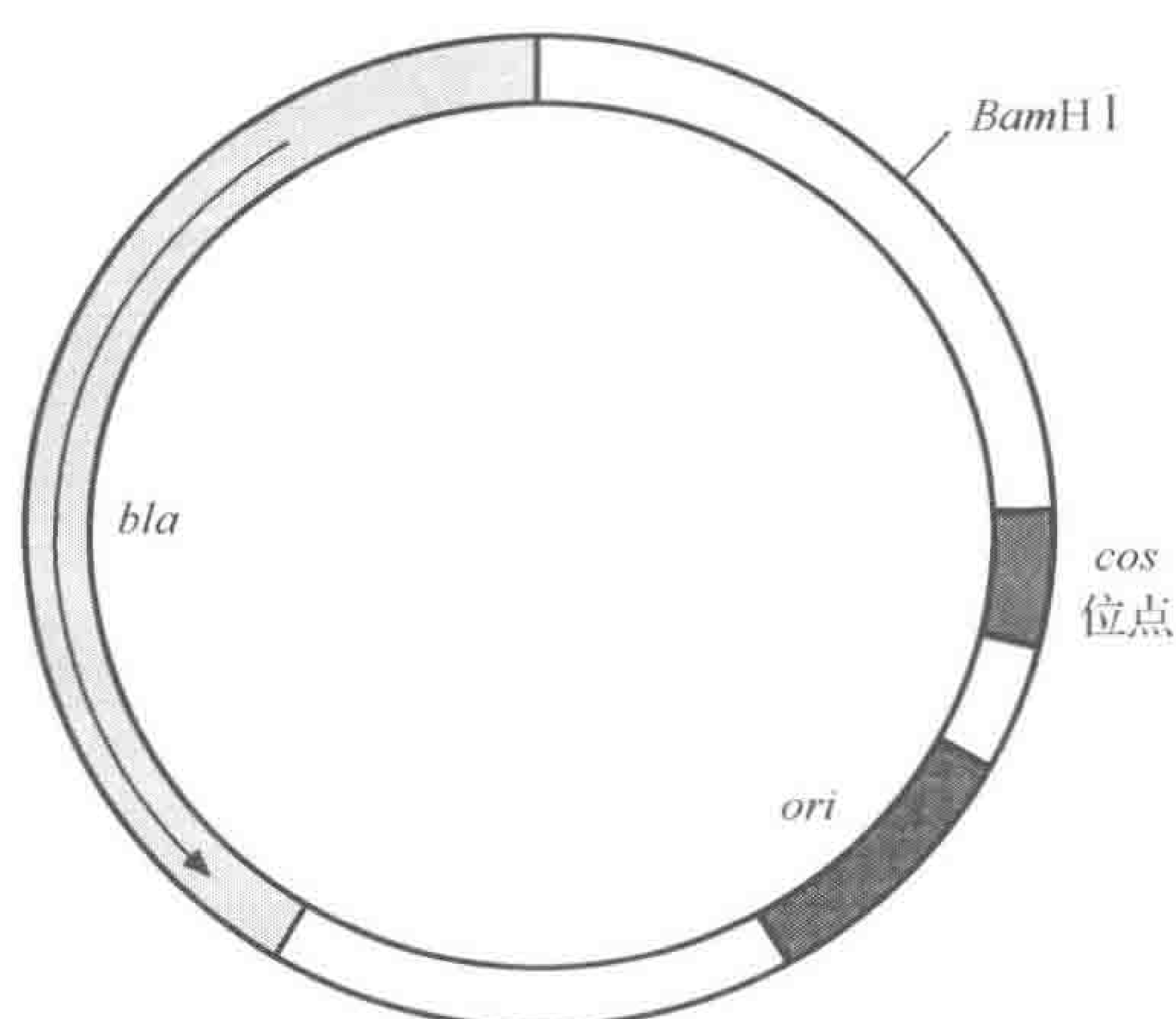


图 8.8　黏端质粒的结构

ori，复制的起点；*bla*，β-内酰胺酶（氨苄青霉素抗性，作为一个选择性标记）；*Bam*H Ⅰ，限制酶识别位点使 DNA 可以插入。*cos* 位点的存在能够使质粒被包装进入 λ 噬菌体颗粒，提供了一个可以插入足够大 DNA 片段的位点

8.4.5　M13 噬菌体载体

丝状噬菌体 M13（见第 4 章）是另一种类型的载体。M13 噬菌体颗粒中含有单链 DNA，而复制型是 DNA 双链。因此，DNA 片段可以被插入到复制型的 M13 DNA 中。转化进合适的宿主后可以产生噬菌斑，从噬菌斑中可以得到重组的噬菌体。因为噬菌体颗粒中含有单链 DNA，这样很容易就获得克隆 DNA 的单链，这在如点突变等应用中很有意义。基因测序最开始只能通过 M13 载体获得单链模板。之后随着技术的发展，可以利用双链作为模板，使 M13 变得不重要，更新的基因组测序方法完全不必使用克隆技术（第 10 章）。

8.5　基因文库

一个基因文库（或者更精确地说一个基因组文库，以便和后文的 cDNA 文库相区别）是一个重组克隆的集合，每一个克隆带有不同的来自于感兴趣生物的基因组的片段，它们代表了生物的整个基因组。发现所需要的基因依赖于对文库进行筛选，并获得所需的克隆。

尽管分子生物学上的很多发展依赖于基因文库的使用，但基因组测序的出现在很多方面提供了另一种更方便的选择。如果生物体基因组已经被测序，只需要简单地使用 PCR 进行扩增，再克隆，就可以得到需要的基因。即使选择的生物体基因组没有被测序，现在也极有可能快速地获得足够完整的基因组序列，筛选和扩增目标基因。不过，基因组文库仍然有它的用处，下面介绍基因组文库的构建和筛选。

8.5.1 基因组文库的构建

构建一个基因组文库包括供体 DNA 随机的片段化和将这些片段的混合物与 DNA 载体连接的过程(图 8.9)。这样将会产生大量的重组分子,每个重组分子都带有来自于供体 DNA 的不同片段。不用试图分离这些分子,相反,要将所有混合物都转化入大肠杆菌。每个克隆都带有不同的 DNA 片段,所有的克隆组成了基因文库。

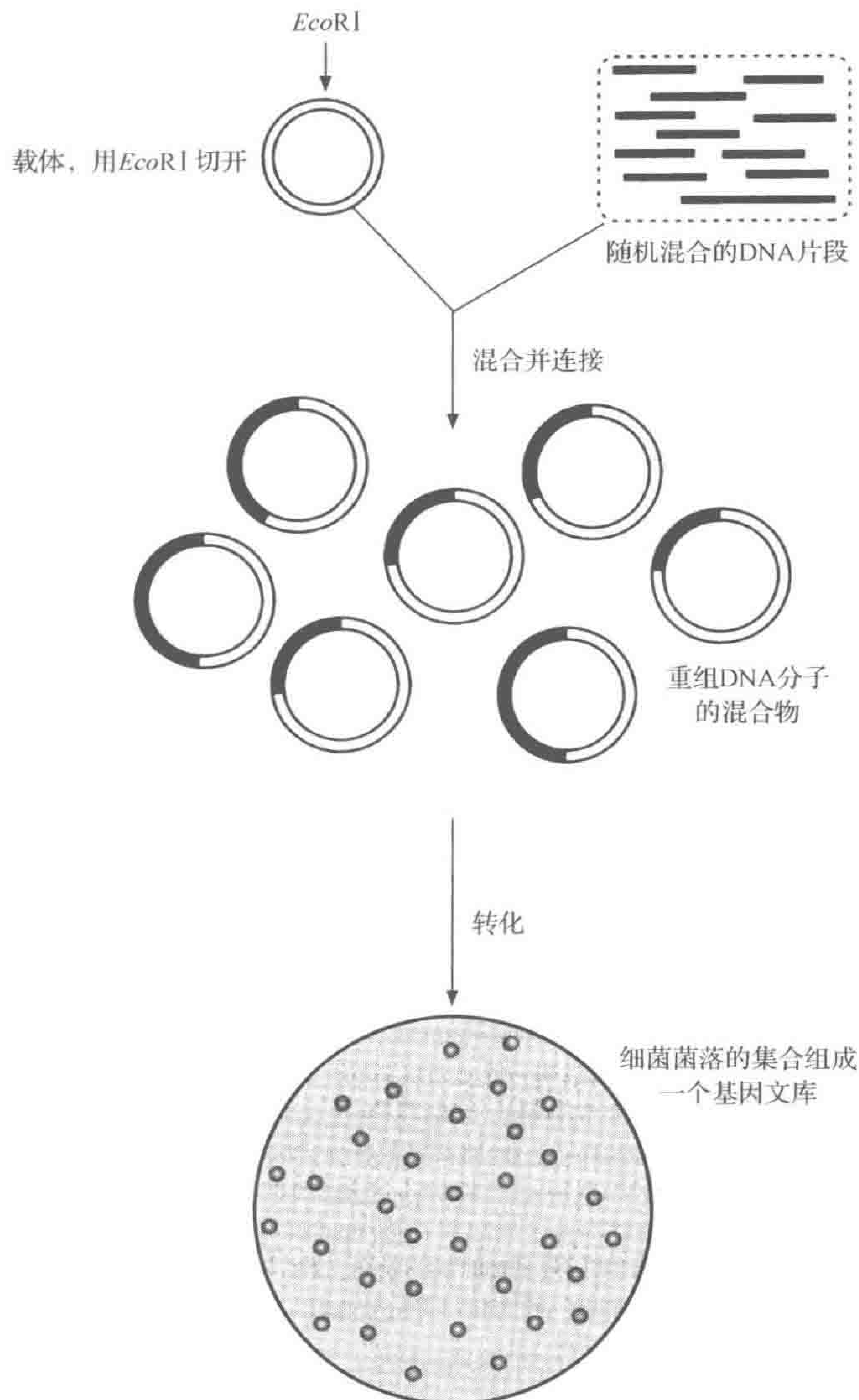


图 8.9 用一个质粒载体建立一个基因文库

一个完整基因组文库所需的克隆数取决于基因组大小和克隆片段的平均长度。例如,对于一个基因组大小为 4×10^6 bp 的细菌,插入片段平均长度为 18kb(使用 λ 噬菌体),有 1000 个克隆的基因组文库有 99%可能包含任何特定基因。如果插入片段平均长

度更小，如 1kb，对于同样的基因组，需要大约 18 000 个克隆才能达到 99%的概率。（计算已经考虑了在随机文库中会出现的重叠和重复）运用下一部分介绍的技术，筛选数千个克隆的文库不是难事，因此，即使是小插入片段的文库对细菌基因组也会非常有用，但对于更大的基因组，更长的插入片段变得更重要。

8.5.2 基因文库的筛选

构建一个细菌基因组文库相对比较简单，然而，需要通过某种方法寻找感兴趣的基因。目前已建立起来的一种快速筛选文库的方法，使用抗体或基因探针就可同时检测文库中的大量菌落。

利用探针筛选基因文库

基因探针的使用依赖于互补的单链核苷酸序列之间的杂交（见第 1 章）。先将菌落克隆（或噬菌斑）转印到滤膜上，然后用标记的基因探针和滤膜进行杂交。目前，非同位素标记的探针得到普遍使用，而非放射性标记。滤膜上呈阳性反应的位点很可能是一个特异性的克隆位点，可以挑取其并进行纯化。

筛选文库的关键因素是怎样获得所需要的探针。如果知道或猜想想要的目标基因与来源于其他物种且已鉴定的基因存在相关性，那么可以以相关序列作为异源探针，进行低严谨杂交（允许 DNA 序列之间一定程度的错配）。如果能够分离获得少量的纯化蛋白，就可能确定该蛋白的部分氨基酸序列，通过氨基酸序列信息，进而推测出 DNA 序列并合成与之互补的短寡核苷酸探针。

另一种获得用于筛选基因文库的 DNA 探针的方法是 PCR（聚合酶链反应）（见第 2 章）。PCR 引物的设计基于数据库中一些相关基因的序列，这些序列通常是基因家族中已知的且非常相似的序列（保守序列）。如果目标基因也包含有相似的序列，那么设计的引物就可与之退火并从靶基因中扩增出一段 DNA 片段（图 8.10），该产物被标记之后就可用作筛选基因文库。

利用抗体筛选基因文库

如果能制备抗感兴趣蛋白的抗体，则可以像使用基因探针一样利用这些抗体筛选基因文库，但这种方法具有一定的局限性：目标基因必须被表达（不必以活性形式），抗体应该能够识别处于变性状态的该蛋白（因此应该识别线性结构的抗原决定簇，而非空间结构的抗原决定簇），而且，该抗体必须识别初级翻译产物而非修饰产物（如糖基化修饰的抗原表位），因为在重组细菌中并不会发生这种转录后修饰。尽管抗体筛选存在着局限性，但还是被广泛应用，尤其在筛选一些特殊病原体关键抗原基因时（见下文 λgt11 的描述）。

通过互补性筛选基因文库

也可通过检测文库中基因与大肠杆菌突变株特异互补的能力来筛选一个基因文库。例如，通过来自于其他细菌携带 *leu* 基因的重组质粒将有可能使亮氨酸缺陷型突变体（在缺乏亮氨酸条件下不能生长）转变为野生型，这样就可以通过将文库接种到缺乏亮氨酸的培养平板上，鉴定所需要的克隆。

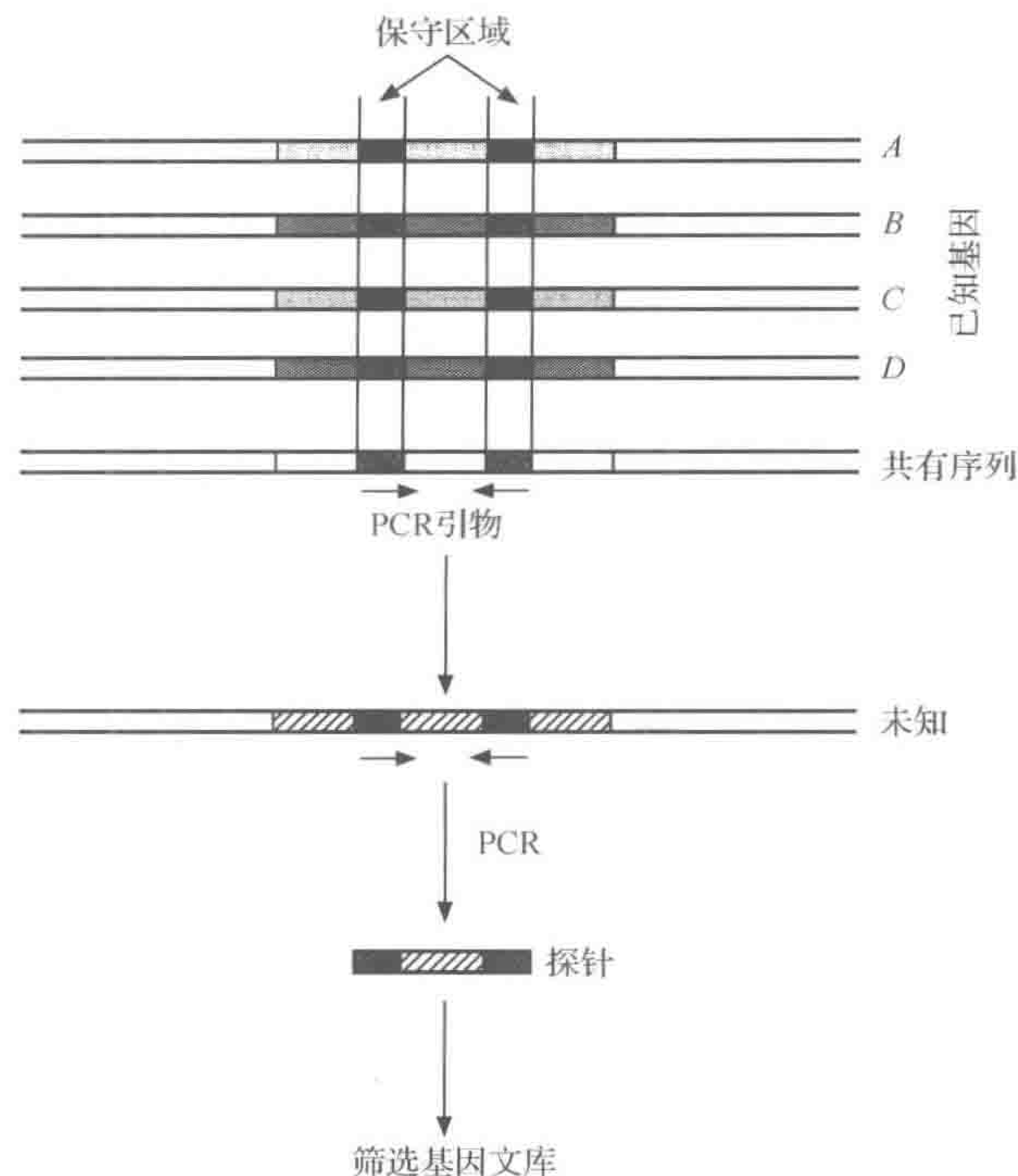


图 8.10 利用 PCR 来获得筛选基因文库的探针

来自不同物种的相关基因 $A \sim D$，都具有两个保守区域(在 4 种基因中具有相似性)。识别这些保守序列的 PCR 引物可以用来从另一物种中扩增出相对应的基因。扩增产物则可用作筛选基因文库的探针，以分离出完整的基因

当然，克隆的基因只有在大肠杆菌中表达且行使其功能时(而且容易检测)，这种方法才适用，因此通过互补性筛选基因文库并不是最佳筛选基因文库的方法。然而，通过序列比对，当已了解一个基因可能的功能时，这种筛选方法对该基因特性的证实将非常有用。

8.5.3 PCR 产物的克隆

利用 PCR 技术可以合成探针，对基因文库进行筛选，然而也可不必构建基因文库，而是利用 PCR 直接克隆目标基因。如果基因组序列已知，则能够确定所感兴趣的基因序列并从它的邻近序列中寻找引物。这些引物可以用来扩增包含有目标基因序列的区域，扩增产物就可用于后续的克隆。如果基因组序列未知，也可通过其他相关物种的基因组序列获得需要的引物，或者简单地测定该物种的基因组序列将在第 10 章中进一步讨论 PCR 在相关物种基因的比较研究)。

回顾最初关于克隆描述的文献将想到克隆 PCR 产物遇到的一个难题，即 PCR 产物不具有对连接反应有帮助的黏端。目前对此问题有好几种解决方法，一种就是在 PCR 产物末端添加酶切位点，这需要在设计引物时在其两端添加酶切位点序列。由于只有引物的 3′端序列能与目标序列互补，因此可对引物的 5′端进行修饰让其带上一个不同的序列，如限制酶切位点(限制位点)(图 8.11)，引物的这部分序列不会与靶序列复性，但影响轻微，只要引物的 3′端能与目标序列复性，且 5′端非复性尾巴不是很长，扩增反应仍能顺

利进行，由此扩增出的产物能够被合适的限制酶切割产生黏端以利于后续的连接反应。该方法不仅能用于基因的初始克隆，它也是亚克隆基因的一种极其有效的方法，如将一个基因从一个载体转移到另一个载体（如表达载体）时。





图 8.11 使用修饰的引物扩增 PCR 产物
在引物中添加限制位点有利于产物的克隆

克隆 PCR 产物的另一种方法是利用 Taq 聚合酶的末端转移酶活性，它可以在 PCR 产物的 3′端额外添加单个脱氧腺苷（A），而克隆载体在线性化后将带上一个悬挂的 3′脱氧胸苷（T）残基，该残基可与 PCR 产物的脱氧腺苷残基互补，因此不需要任何修饰就能完成连接反应。

8.5.4 cDNA 文库的构建

在某些情况下，以 mRNA 而非基因组 DNA 为来源构建基因文库显得更为容易，但这需要以 mRNA 为模板合成双链 DNA。在这个过程中可能存在很多可变因素，但并不需要过多考虑，最重要的是如何使用反转录酶（它是一种 RNA 依赖的 DNA 聚合酶，与此酶相反的负责严格意义上转录的酶是 DNA 依赖的 RNA 聚合酶）。合成的 DNA（互补的 DNA 或 cDNA）可连接于载体用来构建 cDNA 文库。

cDNA 文库普遍适用于动物和其他真核生物，因为 cDNA 文库仅包含在起始材料中表达的那些基因，这远比基因组文库小得多。而且，真核基因通常包含内含子（见第 1 章），这些内含子通常在 mRNA 加工的过程中已被除掉，所以包含全部基因的 cDNA 克隆比相应的基因组 DNA 要小的多。

细菌 cDNA 文库并不常用，因为细菌基因组比较小，且通常不含内含子，因此构建细菌 cDNA 文库没有很大优势（况且细菌基因组文库很容易制备）。尽管如此，在分离和鉴定一些在特殊生长条件下选择性表达的基因时，细菌 cDNA 文库仍然很有用。虽然现在微阵列以及转录物快速测序的应用逐渐使以此为目标的 cDNA 文库的使用被淘汰，作者将在第 10 章讨论反转录酶的其他应用。

8.6 克隆基因的表达

8.6.1 表达载体

如果要把一个基因从一个生物体放进另一个生物体，假如这种情况存在，它可能也不会很好地表达，通常的原因是新宿主中的 RNA 聚合酶不能识别原来的启动子。所以如果要成功表达该基因，就需要使用包含有与克隆位点相邻的大肠杆菌启动子的表达载体（图 8.12）。在克隆位点以正确方向插入 DNA 片段，并位于载体启动子的控制之下，以确保该基因在重组的大肠杆菌克隆中成功地表达。

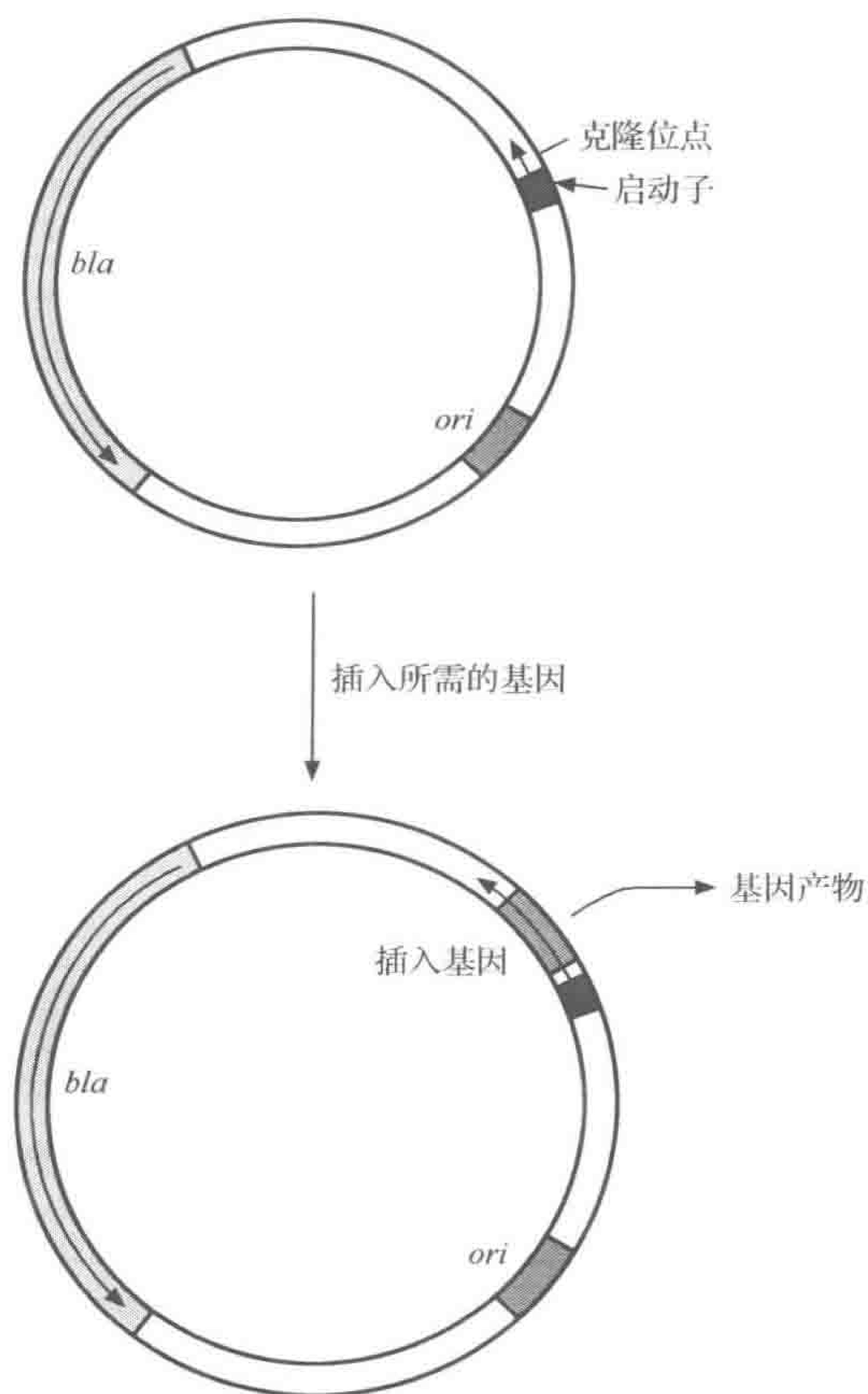


图 8.12 表达载体的基本特点和使用

载体上克隆位点融合了一个启动子，所以插入的 DNA 能够在该启动子下进行转录。某些表达载体也含有翻译信号

对载体的进一步改造和运用则是使其不但能够提供启动子而且可以提供翻译信号，结果（如果你设计的结构包括一个正确的阅读框架）表达的融合蛋白不仅具有由载体提供的 N 端序列，还具有要检测的 C 端区域，这将提高蛋白质的表达效率并增加其稳定性。

通常，一旦目标蛋白在大肠杆菌中表达，下一步即是对蛋白进行纯化并将其与细胞内其他的蛋白分离。可以使用特殊的表达载体简化纯化步骤，这种表达载体通常在目标蛋

白的N端序列整合了一个亲和标记,如具有短的组氨酸残基延伸的肽能够与镍特异地结合。如果控制这个目标基因以便目标蛋白在它的N端带上一个组氨酸标签(通常为6~10个组氨酸),这种组氨酸标记蛋白就能在结合有镍离子的固相柱子上得以纯化。

λgt11(表达载体噬菌体DNA λgt11)

尽管与其他载体的使用有些不同,噬菌体DNA λgt11可作为一种表达载体的例子,(图8.13)。它是一种在β-半乳糖苷酶基因中插入了单一*Eco*RⅠ酶切位点的插入载体,插入的DNA片段将位于乳糖启动子的控制之下。而且,如果DNA插入的方向和阅读框正确,那么编码的肽段将与β-半乳糖苷酶相连。这种载体在用抗体筛选的基因文库中非常有用,如在鉴别和分离一些编码病原体抗原蛋白的基因时。

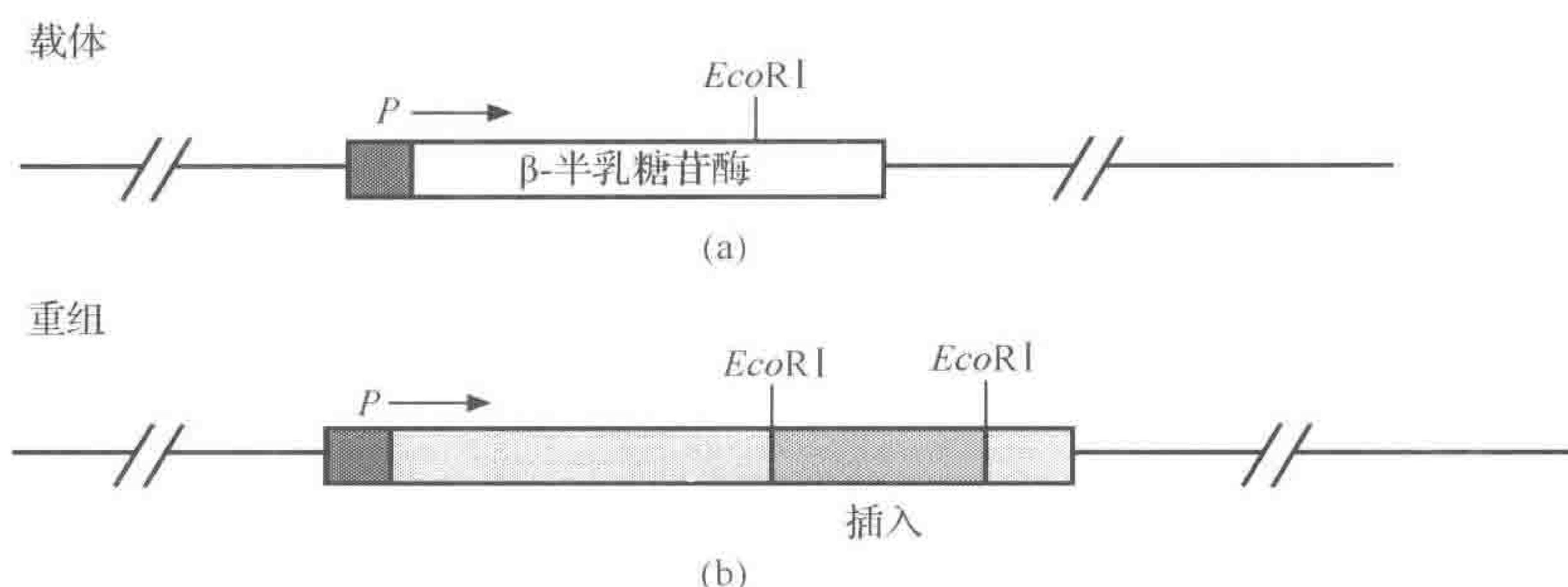


图8.13 使用λgt11生产融合蛋白

(a)载体携带一个β-半乳糖苷酶基因,在该基因中包含了一个*Eco*RⅠ酶切位点。(b)*Eco*RⅠ酶切位点DNA的插入使β-半乳糖苷酶基因断裂。载体不能表达功能性的β-半乳糖苷酶,但可表达一个融合蛋白

产物形成的最大化

基因克隆的最早应用就是通过微生物宿主来生产外源蛋白,例如,人类生长激素(HGH或生长激素)是一种用于治疗脑垂体侏儒症的多肽荷尔蒙,在基因克隆出现以前,生长激素的唯一来源是由尸体解剖分离出的脑垂体。由于供应有限,加之潜在的感染危险,这就促使人们寻找另一种生产生长激素的来源。在大肠杆菌中,克隆表达该基因为大量生产这种荷尔蒙提供了一种极其安全有效的方法。

为了充分利用在细菌中表达外源基因的优点,采用超级表达载体使重组表达产物的量达到最大化。一些高表达的特殊蛋白,使用超级表达载体,其产量可达细胞总蛋白质的50%。然而细胞要将如此多的原料和能量用于合成蛋白质产物,除非它减慢自身的生长速率。(如第5章所提及,这将增加质粒的不稳定性)。通常,解决这一问题的方法是启动子被阻遏,在通过改变生长条件诱导产物形成以前,使细菌持续生长能够达到一个较高的细胞密度。更多的资料和例子见本书附录A。

8.6.2 新基因的获得

DNA合成

基因克隆并不受限于在自然界中存在的DNA序列,可以使用DNA合成仪合成任意

序列的 DNA 片段(寡聚核苷酸)，只需在合成器上输入所需的 DNA 序列，其余的都由机器来完成。虽然这是一个很有序的过程，但随着寡聚核苷酸的延长，精确度会随之降低，但合成碱基数少于 100 的核苷酸仍可以保持较高的精确度。还可以通过酶化学的方法将这些短的寡聚体拼接，从而获得较长的 DNA 序列。

定点突变

可以用合成的 DNA 制造一个完整的新基因。实际上，通常是对天然基因进行诱变，即定点突变。定点突变的方法有很多种，其中一种方法的基本思想如图 8.14 所示。

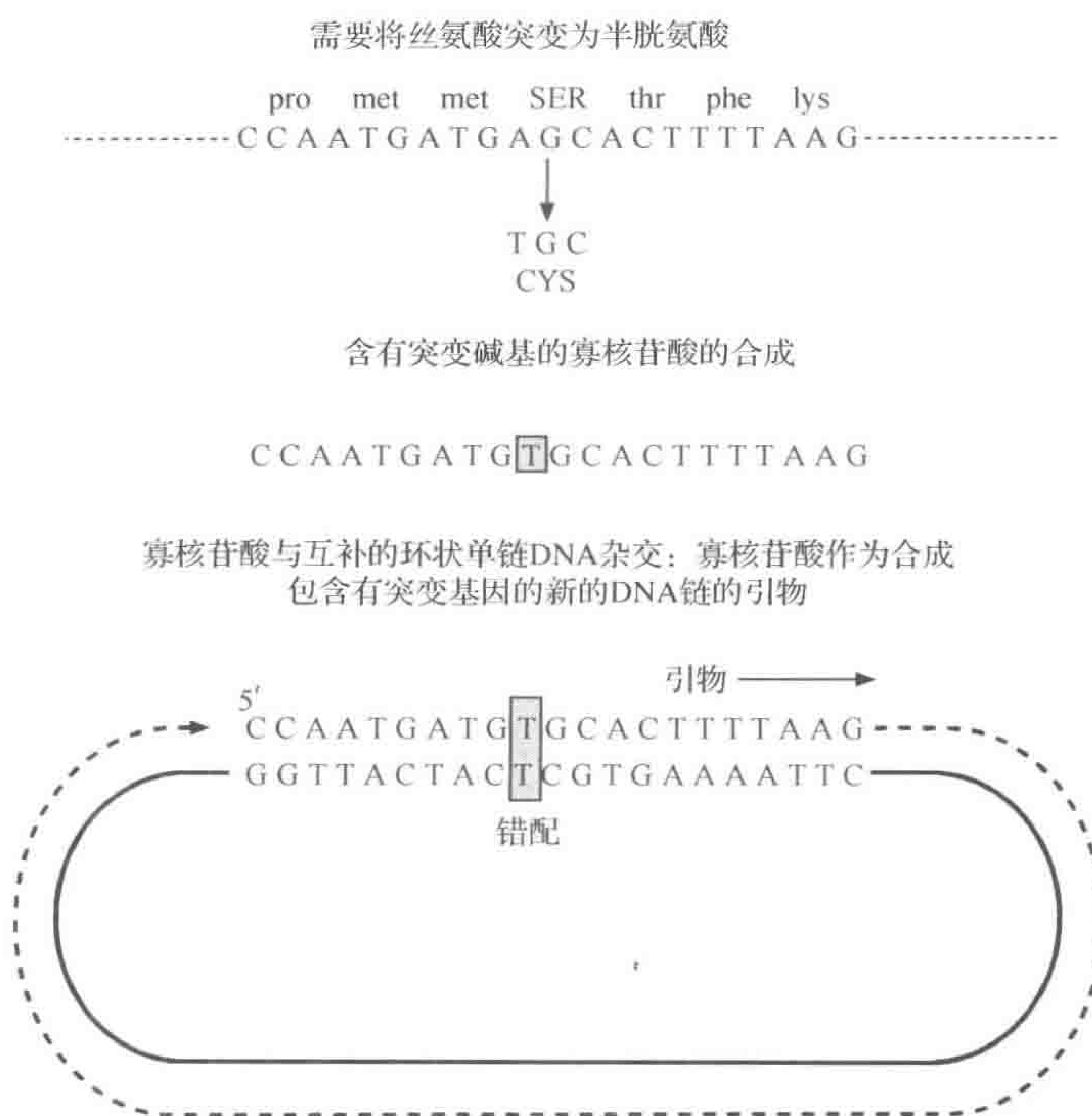


图 8.14 定点突变

该图显示了定点突变的基本思想，还有多种不同的方法可以实现定点突变

首先，合成含有突变序列的寡核苷酸，然后，使之与含有互补目标基因的质粒杂交，尽管这两段序列并不完全一致，但它们足够相似以至于能相互结合，这样，合成的寡核苷酸(错配引物)就能引导完整 DNA 链的合成，并使之带上突变的序列。质粒在转化之后，突变的序列也将获得复制，所以一部分细胞将含有新的基因，而另一部分细胞仍然包含旧的基因。可以用错配的引物探针对这些基因进行区分鉴定。如图 8.14 所示，β-半乳糖苷酶活性位点中的 Ser(丝氨酸残基)在青霉素内酰胺键的水解作用中发挥重要作用，如果将丝氨酸突变为半胱氨酸，将产生一种具有某些特殊性质的新型酶。

蛋白质工程

定点突变的一个应用就是生产一些具有新特点的蛋白质。例如，一些商业化的酶因热不稳定性，其使用常常受到限制。如果这种酶很稳定，那么就可以在一个较高的温度下

进行催化反应，这样就可增加反应速率和整体催化效率。另外，底物结合位点周围氨基酸序列的微小改变都将影响酶对底物的识别，而用传统的遗传学方法实现这种类型的突变并不容易。

定点突变技术及其许多在此基础上发展起来的技术为蛋白质工程新兴领域的研究奠定了基础。然而，蛋白质肽链中任一氨基酸的替换都可能对其结构产生不利的影响，因此，合理应用蛋白质工程，不仅要详尽地了解蛋白质晶体细微结构，而且要熟知什么决定其构象，以及酶的结构与功能之间的关系。迄今，对其理解仍然具有片面性。

蛋白质功能的体外定向进化

在上部分内容中，介绍了在基因中引入特异性突变的方法，用来生产具有新功能或改进特征的酶类。然而，定点突变技术的应用不仅受到对蛋白质结构和功能了解有限的限制，而且，如果想对该基因作大幅度修饰时，也将费时费力。另一种技术，体外同源重组技术，能利用 PCR 的方法在试管里通过重组实现蛋白质体外快速和直接的进化。在这个过程中，来自不同物种的相关基因或具有相关功能的基因被片段化，然后将它们混合，并通过 PCR 使之随机组合(装配)。新组合的基因被克隆进大肠杆菌，以鉴定哪些基因能表达出需要(感兴趣)且有用的产物，例如，通过对这些克隆子的筛选来找到一种降解新底物的酶类。体外同源重组的概念也可扩展为将具有随机序列的合成的 DNA 片段取代该基因的特定部分。

在某种程度上，体外同源重组技术回归到了之前的研究思路，就是挑选具有某些特性的随机产生的菌株，而不是通常与基因克隆技术相关的更特殊的方法，但是用这种方法能够加速新突变体的产生。

8.6.3 其他的细菌宿主

尽管大肠杆菌是基因克隆最通用的宿主细胞，但它还是具有自身的局限性。一方面是难以获得产物分泌物，蛋白质分泌物对商业生产尤其有用，因为它们能加速蛋白质产物的纯化。此外，蛋白质在细胞内的高水平表达通常会导致产物的不溶性，因此，表达的蛋白质在细胞内很容易形成不溶的聚集体。这些聚集体一方面对细胞不利，另一方面也难以重新溶解。细胞外培养基容量很大，所以对于分泌型蛋白而言不存在上述问题。使用其他细菌宿主，如枯草芽孢杆菌(*B. subtilis*)，更易于蛋白的分泌。

细菌的遗传修饰并不局限于生产外源蛋白。一种细菌的特定遗传结构的特异性改变可以提供关于其生化生理过程以及其他重要特性的难以估价的信息，尤其是对其病原性的理解。这包括载体的发展和对于许多细菌进行修饰的其他技术。

穿梭载体

使基因克隆及相关技术应用于各种细菌的一个很大难题就是可用的技术对这些生物体是受限的。例如，转化(电转化)效率可能很低，筛选基因文库的方法也将非常困难。因此，用大肠杆菌作宿主来实现起始克隆和基因的操作非常有利。然而，当确定获得了少量的正确 DNA，就可以从重组载体上重新获得该 DNA 并导入目标细菌。这个过程可由一

种能在两种细菌中复制的载体轻松完成。这种载体可以通过在大肠杆菌质粒上插入一个复制原点而获得，且该复制原点可以在第二种宿主中行使功能，于是该质粒具备了能在两种细菌中复制的能力。由于这种载体能在两种细菌中携带 DNA 片段，因此称其为穿梭载体(图 8.15)。广谱质粒也可作为一种二元复制起点的穿梭载体。

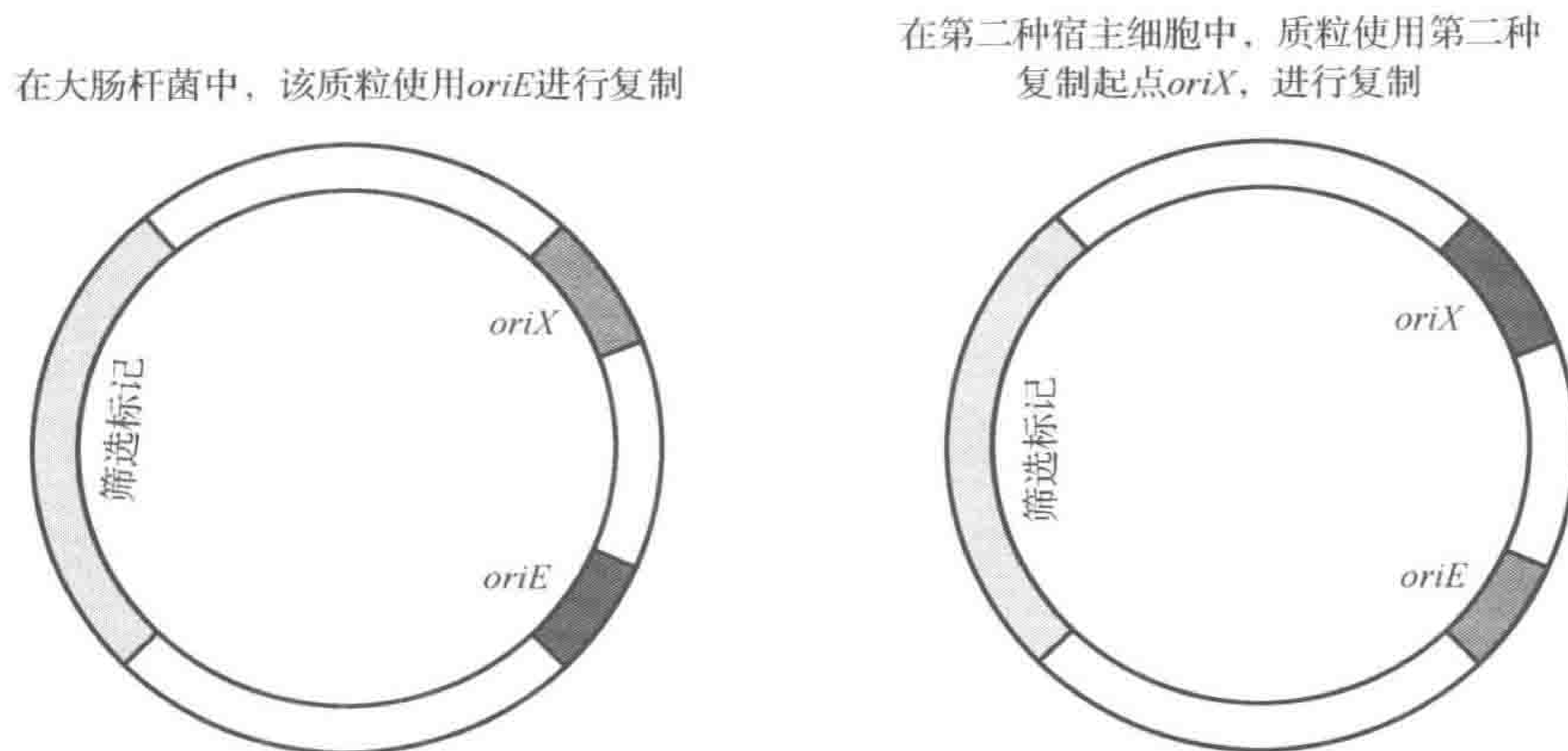


图 8.15 穿梭质粒

原生质体转化

如前所述，适用于大肠杆菌的转化系统，并不容易适用于其他宿主菌的转化，这就需要其他的转化系统，其中一种就是原生质体转化系统。酶法除去细胞壁，在渗透压稳定剂(如甘油)的存在条件下，形成仅由细胞膜包裹的原生质体，然后加入 DNA 溶液，在聚乙二醇的介导下使细胞摄取 DNA，之后除去聚乙二醇，并使细胞在渗透压稳定的培养基中再生。如果转化条件合适，将获得高比例的转化克隆。该方法可应用于多种细菌(如链霉菌属和链球菌属)的转化。

电穿孔

尽管原生质体转化适用于多种细菌，但是对于不同物种，转化效率也不尽相同。而且有时转化效率太低以至于不能用于实际应用，另外，原生质体形成和再生条件的优化也比较困难。电穿孔提供了更通用的一种选择，在这个过程中，细菌细胞与质粒 DNA 混合后，对其施加瞬间高压脉冲，细胞膜上将形成瞬时的微孔，通过微孔 DNA 进入细胞，这种重组转化子可以通过质粒 DNA 上的抗生素基因得以筛选。

尽管在电穿孔条件优化方面，仍需做大量工作，但理论上，该方法能够将 DNA 导入任何一种细胞。

根瘤农杆菌：自然界的遗传工程师

新的科学发现常常是充满活力的科学家在探索事物本质的创新中完成的。一个例子就是，在分子生物学出现很久以前，一种简单的土壤细菌生产转基因植物已有千年了。这种土壤细菌，即根瘤农杆菌(*Agrobacterium tumefaciens*)，它是作物冠瘿病的诱发因素。在这种疾病中，虫瘿实际上是一个植物肿瘤。最令人兴奋的发现是，该细菌 DNA 的一部分特异性片段被转移进植物细胞，并整合进植物染色体上，从而形成肿瘤。在这个过程

中,一个重要的工具就是根瘤诱导质粒(Ti)(见第 5 章)。这种 DNA 转移现象已被广泛用于生产转基因植物,并为植物分子遗传学奠定了基石。

8.6.4 新疫苗

至今,疫苗可分为 3 种类型:①减毒活疫苗,指通过突变而使其失去病原性的病原体(如口服脊髓灰质炎疫苗);②灭活疫苗(如流感疫苗);③类毒素(如白喉类毒素、破伤风类毒素),指已经失去活性的纯化毒素。现在,越来越多的研究发展增加了新型疫苗生产的可能性,如亚单位疫苗,包括非细胞百日咳疫苗和流感嗜血杆菌乙型疫苗,前者是由百日咳杆氏菌(*Bordetlla pertussis*)的某些细胞成分纯化物组成的,后者则包含了来自流感嗜血杆菌(*H. influenzae*)的结合蛋白载体的多聚糖荚膜。

遗传技术从多方面促进了新疫苗的发展。首先,对于亚单位疫苗而言,克隆一个基因并使之在便利的宿主细胞中高效表达是可能的,这不仅使大量纯化蛋白的生产变得更为容易,而且还能避免由生长的大量病原微生物带来的危害。一个很好的例子就是抗乙型肝炎病毒疫苗的生产。克隆表达乙型肝炎病毒表面抗原能够生产出现在日常应用中更为安全有效的乙肝疫苗。(在此情况下,由于抗原在大肠杆菌中不能以正确形式合成,因此通常使用酿酒酵母来获得乙肝疫苗)。

对于一些疾病,需要使用活疫苗。生产活疫苗的一种方法是在一个现有的疫苗株基因组中插入一个合适的基因,这将产生一种表达外源蛋白的重组疫苗,因此可被作为活疫苗使用。用于开发这种疫苗的主要载体是天花疫苗病毒,即牛痘苗,实际上,在一种病毒中包含来自不同感染因子的基因是有可能的,这就大大增加了接种一个疫苗同时预防多种疾病的可能性。同样,如伤寒沙门氏菌减毒疫苗和结核疫苗——卡介苗(BCG)正被用于活的重组疫苗生产。

利用基因技术生产活疫苗的另一种方法是利用病原菌的合理减毒,在实验过程中,如果能够确定某些基因对于毒力的产生是需要的,但对于生长是不需要,这样就可以通过基因敲除技术(见第 9 章)特异性地使这些致病基因中的一个或多个基因失活,从而生产非致病菌株,因为可以确保相关基因被完全删除且不会有毒力复原的可能性,这要比传统的基因突变更有效。了解病原体全长基因组序列对通过基因替代的方法寻找合适的减毒潜在靶标尤其重要。

基因克隆为疫苗的生产提供了另一种更加新颖的途径。如果克隆一个关键抗原,并将其插入一个在动物细胞中发挥功能的启动子控制之下的质粒载体,那么,这种重组质粒自身就能够引发免疫反应。现在,这种 DNA 疫苗仍停留在实验阶段,其安全性和效能仍有待评价和确定。

8.7 基因技术的其他应用

在本章中,详细探讨了基因技术在细菌中的应用。现在公认的一些基因技术方面的应用包括:遗传病分子诊断,人 DNA 指纹技术,人类疾病的基因治疗以及转基因动植物的生产,这些都在本书的讲解范围之外,附录 A 中将提及相关内容的阅读建议。

9 细菌研究的遗传学方法

遗传学是对细菌的代谢、生理和毒力认识进行深化的核心。在这一章里，将用几个实例来学习遗传学技术的应用。分子生物学方法也可适用于细菌的快速识别以及菌株类型的鉴定，以快速追踪感染暴发的根源。在第 10 章里，将进一步讨论基因组测序及相关技术的巨大进步对这个领域的影响。尽管基因组学对了解细菌起了巨大的促进作用，并且成为现在最受青睐的技术之一，本章所述的技术对了解有关基因在细胞整体行为中的功能依然很重要。另外，作为细菌鉴定及其分型的方法，因为这些遗传技术得出结论的准确性和快速性，它们已经取代一些传统的组织培养和生化技术。

9.1 代谢途径

运用遗传分析来研究新陈代谢途径的最初动力来源于 20 世纪 40 年代 Beadle 和 Tatum的研究工作。虽然他们研究的是真菌——脉孢菌(*Neurospora*)，但很多原理随后都应用到了细菌。作者将运用假设的新陈代谢途径来阐述这一理念，该途径会形成必须产物 P(图 9.1)。营养缺陷型突变体(只有在必须产物 P 的存在下才能生长)对应于 3 个步骤，应该包括 3 种缺陷型突变体，并可通过互补实验分析分配为 3 组。

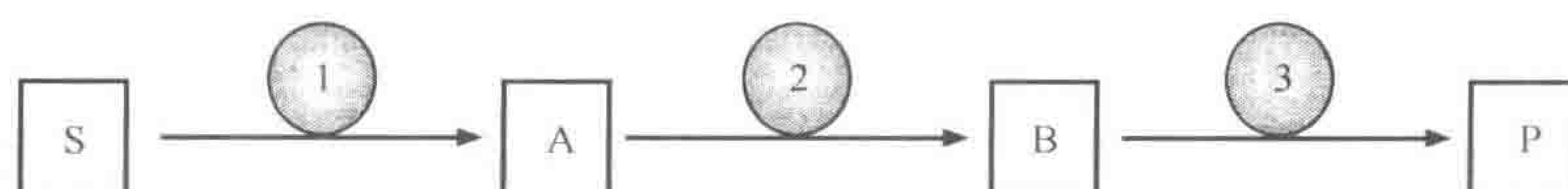


图 9.1 一个未分叉的代谢途径

9.1.1 互补

在第 4 章中，了解到用两种噬菌体突变体进行混合感染可以导致两个突变体之间的互补效应。这一情况是在这两种噬菌体包含不同的突变基因，从而使它们之间含有一套完整的功能基因的条件下实现的。如果把互补分析应用到如图 9.1 这样的突变体途径上，可以把突变体归纳为和它们的缺陷步骤相对应的 3 组(假设每种酶只含有单一的多肽)。例如，缺失基因 1 的突变体之间不能互补，但它们可以弥补另外两个基因任一个基因缺失的突变体。与细菌系统相比(除噬菌体之外)，这种类型的互补分析更容易在脉孢菌上进行。

在细菌系统中，见到的互补效应更多的是与质粒相关的。引入的质粒将在细菌中产生部分双倍体：一个版本的基因在染色体上，另一个版本的基因在质粒上。如果在染色体

上的基因是缺陷型的，其基因功能可以由质粒上的基因互补（第 6 章）。与之相似，质粒互补特定突变体的能力可以用来确定一个克隆基因的功能（第 8 章）。

9.1.2 营养共生

互补分析无法确定参与的步骤及发生的顺序。这些信息是通过营养共生实验来获得的（图 9.2）。没有一个突变体可以在缺乏必需营养成分 P 的培养基上生长。但是，如果照图示的方式进行接种，一种突变体所产生的中间代谢物有可能促进相邻突变体的生长。例如，突变体Ⅰ自己不能产生中间物 A，但突变体Ⅱ可以产生中间物 A。那么，那些接种在突变体Ⅱ附近的突变体Ⅰ就可以被刺激生长起来。类似的，突变体Ⅱ被突变体Ⅲ刺激生长起来。因此，在某种程度上，通过对应的生化途径上的步骤，交互共生实验可以用来对突变体进行分类。

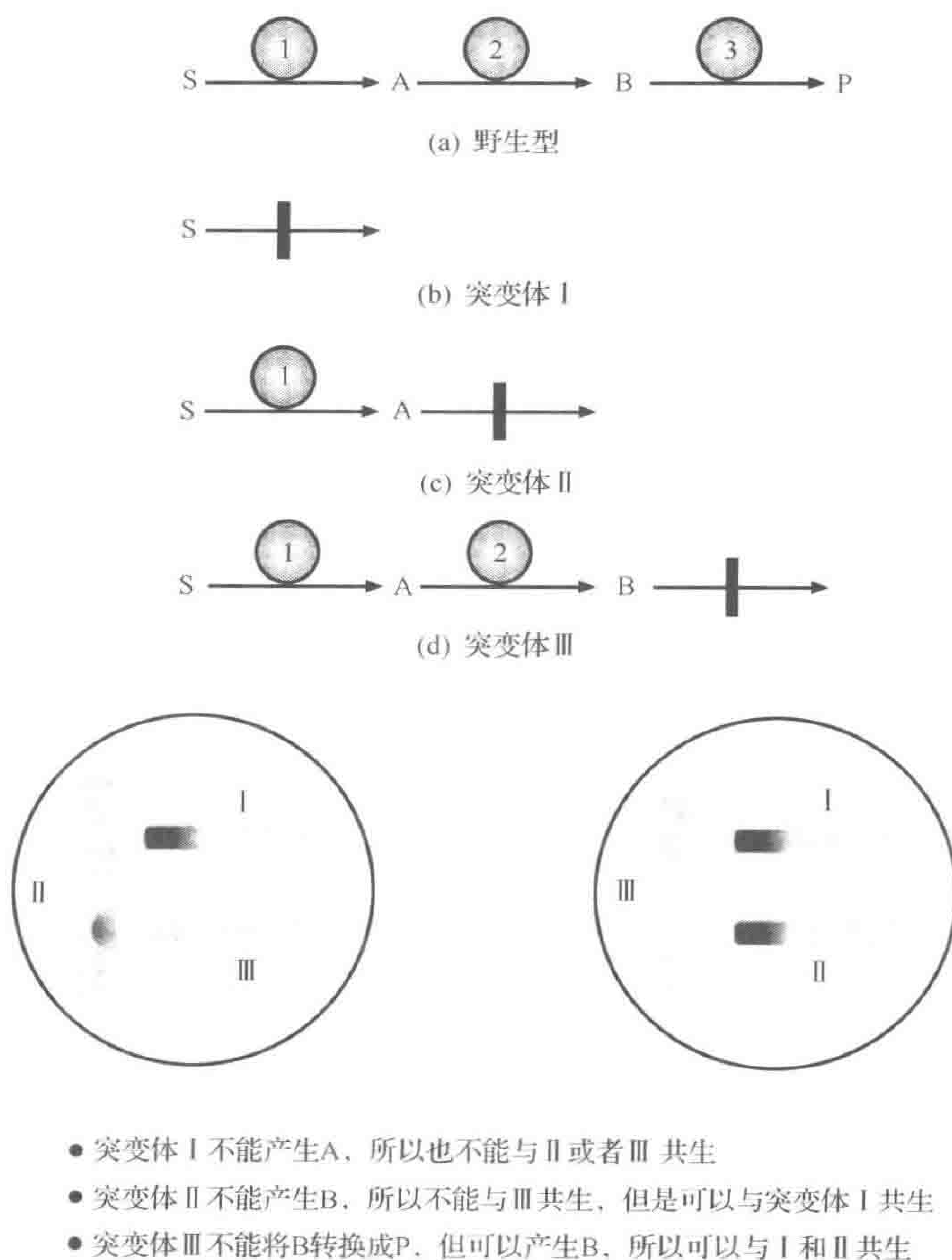


图 9.2 交互营养共生

虽然互补分析和交互共生实验不能告知是哪些步骤或是哪些中间产物参与代谢途径，但是，通过检验预测的中间物对每个突变体生长的影响，可以检验理论假设的代谢途

径。应用这一方法，可以证明鸟氨酸和瓜氨酸是精氨酸生长合成代谢途径的中间产物。第一类突变体只有在精氨酸存在下生长，第二类突变体在有精氨酸或瓜氨酸存在下生长，第三类突变体在有精氨酸、鸟氨酸、瓜氨酸中的一种氨基酸的存在下生长。这证明合成途径是从鸟氨酸到瓜氨酸，再到精氨酸。

9.2 微生物生理学

在上述几个简单代谢途径的例子中，遗传方法学和生物化学实验能够很有效的互补。遗传学很有用，但并不是绝对不可缺少的。但是，一个细菌细胞包含有远比一系列直接的生物化学途径复杂得多的成分。它包括复杂的结构，如鞭毛、核糖体、细胞壁以及像控制细胞复制和分裂的复杂系统。虽然在一些情况下，可以采用还原方法，如核糖体可以在体外分散及再组合。但是，大部分情况下，这是极端困难甚至是不可能的。在研究这种系统时，遗传学发挥了重要作用。

图 9.3 总结了一些分析生理学特征的遗传基础的方法。在这一例子中，分离到在某一特定性状下的系列突变体(如孢子)是研究的起始点。根据它们表型的自然性状，可以

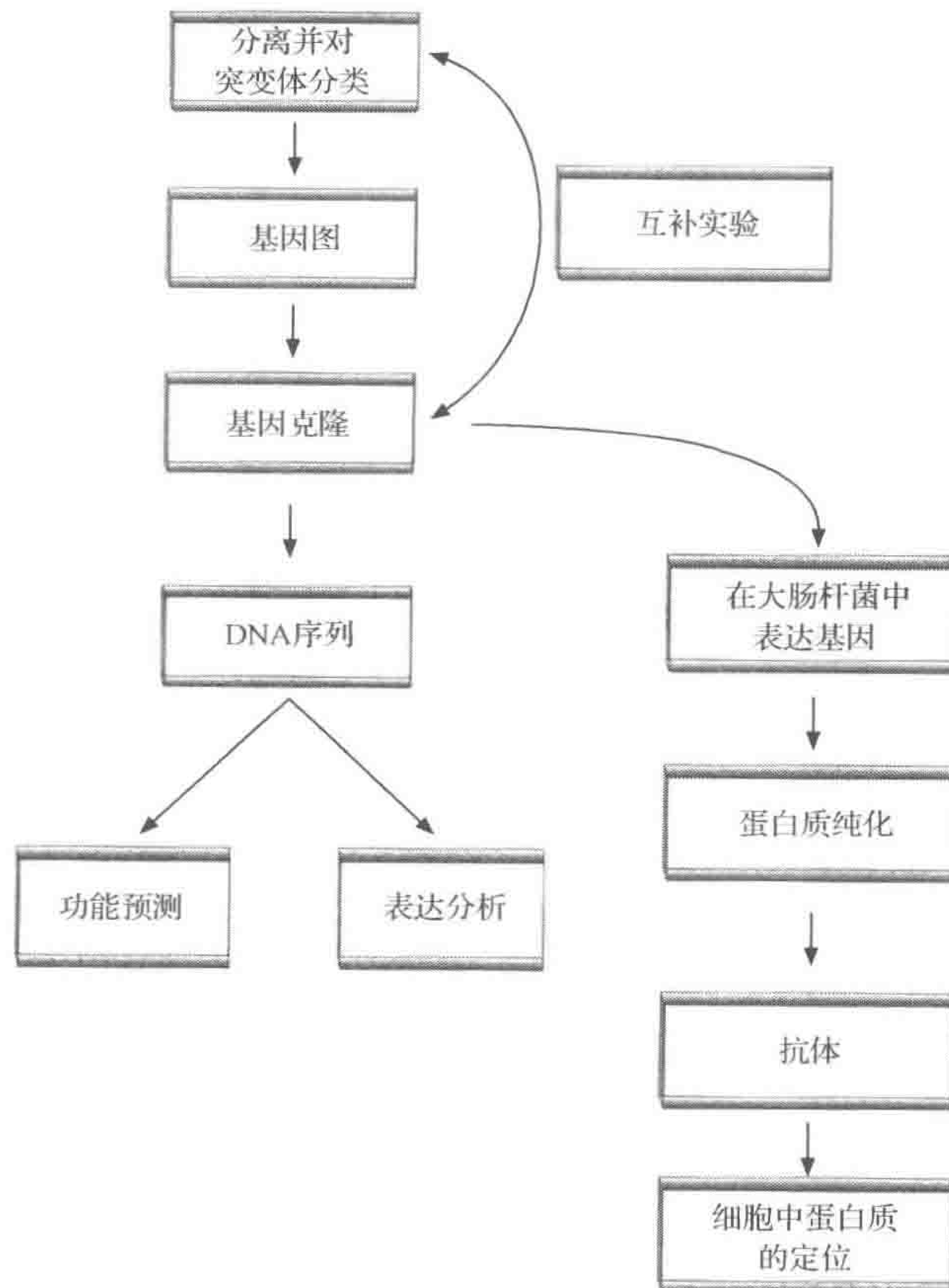


图 9.3 表型特征的基因调查的技术

流程图阐明了某一特殊突变体起始的路径。还有许多其他方法可以进行克隆和识别基因

对这些突变体进行分类，也可由互补分析来分类。所涉及的基因可以用第 10 章所述的方法标示出来(如它们在染色体上的位置)。这一信息可以用来鉴定一个文库中的基因。通过它与原始突变体的互补性，可以鉴定突变体克隆。这个基因可进一步被测序，并通过序列分析来预测它的功能(第 10 章)。这一传统方法用来鉴定突变体显然是费时的，并将很快会被在第 10 章所描述的新的测序方法所替代。例如，现在可以同时对亲本株和突变株进行测序，并通过两组序列的直接对比来确定突变的特性。利用序列的数据，可以设计相应的探针对基因在不同条件下的表达进行分析，并有可能获得对突变体所起的作用的数据。而且，克隆的基因可以被超量表达(第 8 章)，以便获得更多的纯化的蛋白，进一步制备出针对该纯化蛋白的抗体，以便用于确定这个蛋白在细胞上表达的部位。

9.2.1 报道基因

遗传学方法在对细菌生理学的了解上起了无法估量的作用，在看几个这样的例子之前，需要了解一下方法学里的另一种技术。不采用突变体，假设在特定条件下，特定需要的基因在一定的时间可以被选择性表达，将不同的基因在不同条件下的表达进行检测。一个方便并被广泛应用的方法是采用报道基因。这种方法采用把要检测的基因调控区与一个容易被检测到的基因连接起来。这样就可以通过报道基因的表达来了解基因的调控(图 9.4)。例如，利用半乳糖苷酶信号基因和含有显色底物 X-gal 的培养基，那么细菌克隆只有在有启动子被激活并使报道基因表达后才能变成蓝色。



图 9.4 报道基因的使用

上图展示了转录融合，启动子缺乏的 *lacZ* 被一个启动子融合，以激活启动子的活性为特征。转录融合被使用，出于某些目的，如在那些启动子片段用来提供转录信号和蛋白质编码片段的 5′端，形成一个包括报道基因在内的融合蛋白

可以用相似的步骤来鉴定那些在特定刺激下被激活的未知基因。在这种情况下，那些含有启动子的 DNA 随机片段与报道基因连接起来产生一个启动子融合体基因库。这个基因库之后被涂到含 X-gal 底物(假设用 β-半乳糖苷酶信号基因)的培养基上。所感兴趣的克隆是那些起始是白色，但当培养条件变化后变成蓝色的克隆，如将培养皿转移到厌氧培养箱中，这表明启动子对这一新环境起反应，在这个例子中是在厌氧条件下培养。通过这个实验，可以推测在这个启动子控制下表达的基因是当细菌在厌氧条件下生长时才会表达的基因。虽然这一方法已经基本上被能对全部基因表达进行分析的基因芯片技术所代替，但在对热、饥饿、渗透和孢子分化等条件做出反应时，它仍然被广泛用于基因表达分析。

除了β-半乳糖苷酶，还有另外两个报道基因值得关注。萤光素酶的表达导致蓝绿光的产生并可以通过测量光的产生来监测基因的表达。它同样使得宿主生物体生物发光，从而可以利用敏感的影像系统来跟踪萤光素酶标记的细菌在复杂环境及生态系统中的行为，这包括跟踪活动物体内的病原体。另一个有利用价值的报道基因是绿色荧光蛋白，一个从海蜇上发现的具有荧光性并在紫外线条件下产生绿光的蛋白。它的优势是能在原位上检测到而不需要任何酶底物。除了作为报道基因用在基因表达上，它也可以作为融合蛋白表达以便使目标蛋白标记上绿色荧光蛋白。这一技术可用于鉴定被标记目标蛋白在细胞中所处的位置。

在下一章，将学习研究基因表达变化的其他方法，尤其是基因芯片的应用。

9.2.2 染色质免疫沉淀

报道基因（以及在第 10 章中将要讨论的微阵列）是用来观察那些受转录因子调控的基因表达的变化（见第 3 章）。如果一个转录因子被鉴定出来，那么首先要问的一个问题就是什么基因受到这个转录因子的调控。这可以通过用微阵列（见第 10 章）来比较在细胞中含有这个转录因子和突变体中不含有这个转录因子的所有转录的 RNA。这看起来好像很简单，可实际上这种研究结果经常无法提供满意的答案，这是由于实验会产生出很长一串很难解释的候选基因名单。用来确定 DNA 结合蛋白的 DNA 结合位点的染色质免疫沉淀是一个可以解释这个问题的更加有效的方法（图 9.5）。为了实施这一技术，细胞被完全破坏以产生由基因组 DNA、蛋白质和其他细胞成分组成的混合物的细胞裂解液。在这种情况下，因为并没有进行纯化，调节蛋白（及其他类 DNA 结合蛋白）仍附着在靶序列的 DNA 上。为了使下一步更易进行，通过交联剂，如甲醛，可以使蛋白质和 DNA 的结合不可逆。交联反应后，结合有蛋白质的 DNA 将被超声波片段化，以便形成平均长度约为 500bp 的 DNA 片段。为分离和被研究的调节蛋白结合的 DNA，抗转录因子的特异抗体被加入混合物中，并导致所研究的蛋白质-DNA 复合物沉降下来，以使它们能从混合物中分离出来。之后，纯化出来的蛋白质-DNA 复合物将被加热来逆转蛋白质-DNA 复合物的交联反应，以使 DNA 能与蛋白质分离开来。通过将分离到的基因组 DNA 克隆到质粒载体，并对插入的 DNA 测序，可以对分离出的 DNA 片段进行鉴定。由于被沉降的复合物在交联反应时应当含有转录因子蛋白所结合的所有 DNA 位点，应用染色质免疫沉淀提供了鉴别转录因子所有结合位点的有效方法。

染色质免疫沉淀与基因芯片的结合

染色质免疫沉淀的限速步骤是克隆和测序。为使这一过程更有效率，染色质免疫沉淀技术与第 10 章所描述的基因芯片技术的有机结合能回答同样的问题而不需要克隆。染色质免疫沉淀与基因芯片结合的过程，采用了染色质免疫沉淀同样的步骤，但是在 DNA 从蛋白质-DNA 复合物分离出来后，要进行 PCR 反应及荧光标记。标记好的 DNA 片段再滴加到由代表基因组所有基因序列的短单链片段组成的网格式 DNA 芯片。标记的片段将只与它的互补链结合。在 DNA 芯片上，特异位点上所产生的荧光显示出与转录因子相结合的是哪个基因。

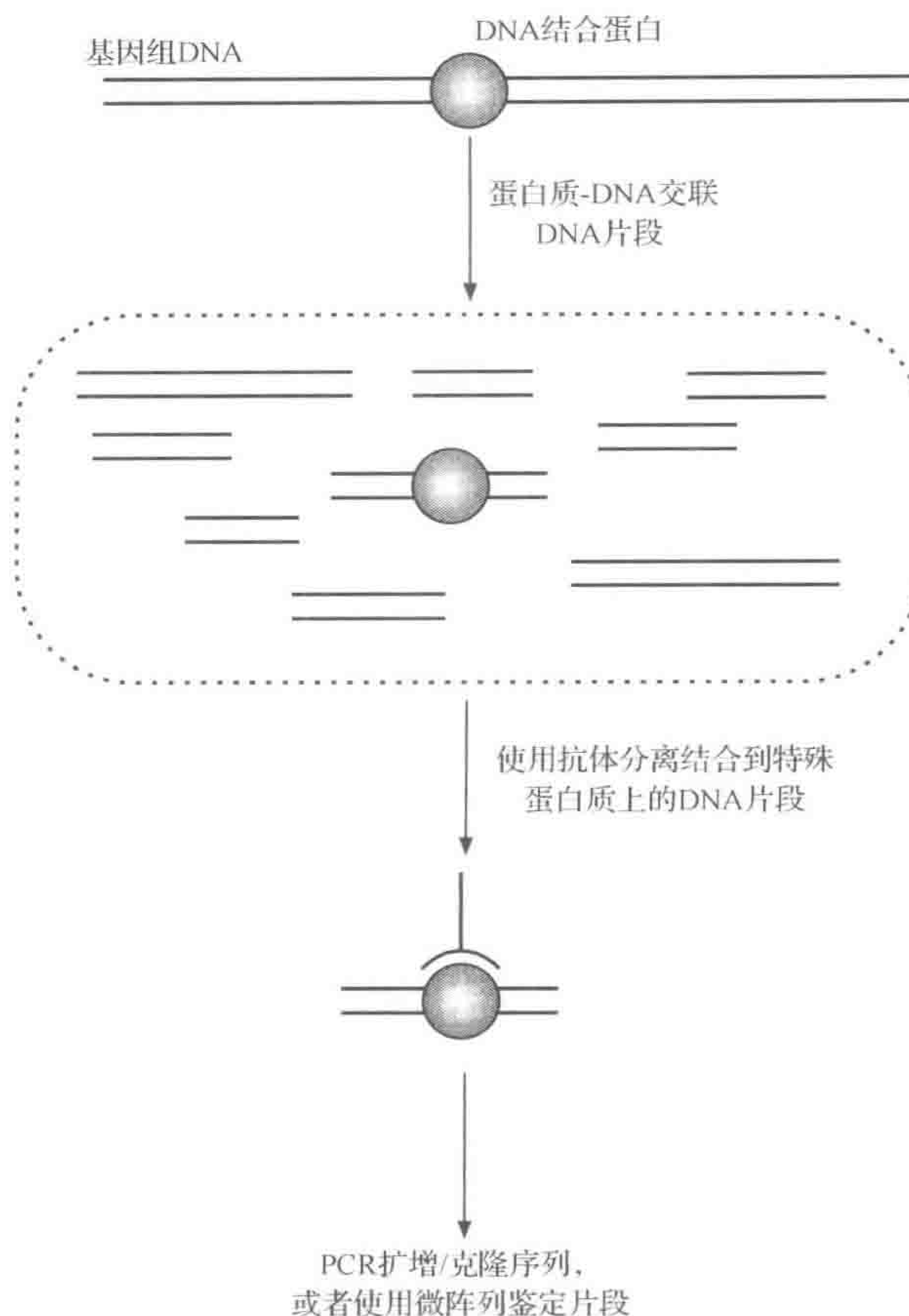


图 9.5 染色质免疫沉淀(ChIP)
使用抗体分离特殊的蛋白质-DNA 复合体，使蛋白质结合的 DNA 序列能够表现出特征

9.2.3 细胞分裂

细胞分裂在细菌生理学上如此重要以至于会让人觉得突变是不可行的，与其相关的突变体将不能存活。令人惊异的是，一些突变体依然能够生长。在这些突变体中，有一种大肠杆菌突变体叫小突变体，之所以这样称呼，是因为在细胞分裂时，一些子代细胞是一些不含有任何染色体的微型细胞(图 9.6)。这些微型细胞当然是不能继续生存下去的，但是因为细胞还能像正常细胞一样进行分裂，突变体还能生长一段时间。

运用条件突变体，尤其是温度敏感突变体来研究细胞分裂也是可行的。细胞分裂失败最常见的后果是形成丝状体，如在杆状型的大肠杆菌中。那些在高温下形成丝状体的突变体被叫做丝状温度敏感突变体。遗传学分析显示，在一组基因中的任一基因产生突变体都会产生这一现象。其中一个相当重要的基因是丝状温度敏感 Z(*ftsZ*)基因。前面介绍的实验方法证明，通过在细胞分裂位点聚合形成环状结构(叫做 Z 环)，FtsZ 蛋白会导致隔离两个细胞的隔膜的形成。

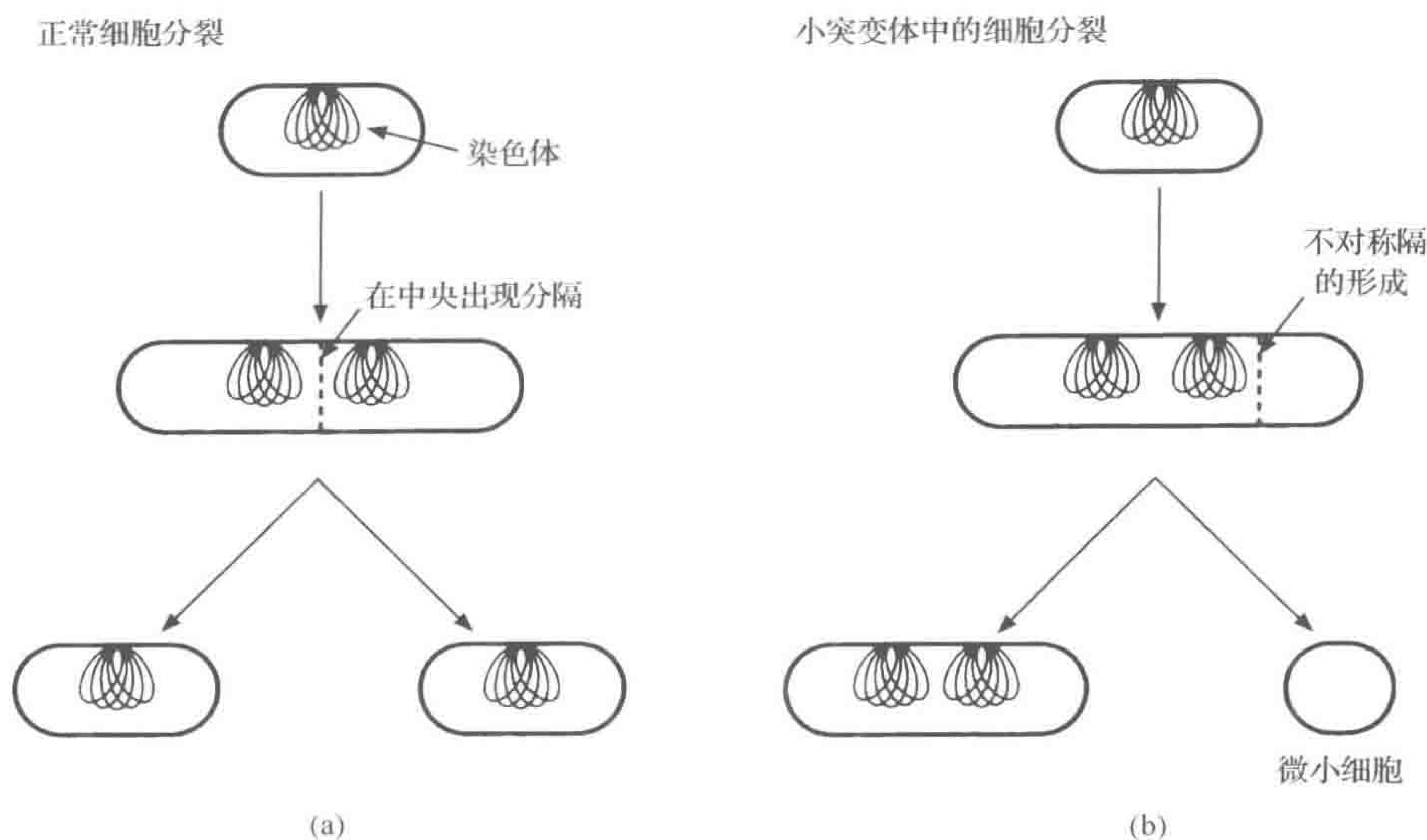


图 9.6 大肠杆菌中的细胞分裂:小细胞突变体
(a) 正常细胞分裂发生在细胞的中央;(b)在小细胞突变体中,
某些细胞分裂是不对称的,产生一个不含染色体 DNA 的小细胞(不复制)

FtsZ 蛋白和小蛋白的功能事实上是息息相关的。大肠杆菌中的第一个微型突变体是在 1967 年分离出来的。引起突变的基因被定位在染色体小 B 位点上,并确定出这组基因编码形成小系统功能的 3 个蛋白质:小蛋白 C,小蛋白 D 和小蛋白 E。这个突变系统对试图解答为什么细胞分裂通常是在中心位点而不是在细胞的其他地方非常有用。完整的答案是相当复杂的,但关键点是小基因蛋白决定了 Z 环形成的位置所在。小蛋白 C 是细胞分裂抑制子。当它激活时,和 FtsZ 蛋白相互作用,抑制 FtsZ 蛋白形成多聚体,从而阻止 Z 环的形成。小蛋白 D 是能直接和小蛋白 C 结合并激活小蛋白 C 的膜蛋白,而小蛋白 E 可与小蛋白 D 结合,从而阻止小蛋白 D、激活小蛋白 C。小蛋白 C 在细胞中的分布状况,及其被小蛋白 D 和小蛋白 E 差异调节下的活力决定了 Z 环形成的位点及随后发生的细胞分裂。

9.2.4 移动性和趋化性

在没有化学因子或其他刺激因子存在的条件下,包括大肠杆菌在内的很多细菌的移动是很随机的单方向平滑移动和剧烈转动的混合方式。这与细菌的结构及鞭毛所控制的局域性转动有关。当转动是逆时针时,丝状细菌形成一束,促使细胞向一个特定的方向移动。而剧烈转动是由鞭毛改变转动方向造成的。一旦鞭毛顺时针移动,团在一起的鞭毛舒展开,细胞暂时转动后,鞭毛又继续它们正常的逆时针转动。细菌就会平缓地向不同的方向移动。

细菌转动频率控制着趋化性(细菌对特异化学刺激的靠近和远离能力)。当细胞向引诱物移动,并且引诱物浓度增加时,平滑移动时间增加(或转动频率降低)。因此,那些移

向正常方向的细菌会平滑移动很长一段后再调整方向，而那些移向错误方向的细菌会及早转换方向。

通过遗传学分析，对这一复杂而新奇的系统有了深入的了解。通过分离出的特异突变体，50多种基因被鉴定出来。这些基因包括①产生鞭毛（组成鞭毛的蛋白质亚基）或组装鞭毛的缺陷型（两种情况都造成无鞭毛细胞，以致细胞不能移动）；②另一类鞭毛缺陷型，导致鞭毛无法转动（这类细胞也不能移动）；③控制转动缺陷型，这导致细胞不停地转动或极少转动，或者失去趋化性，以致细胞可以正常移动，但并不能对刺激因子起反应；④特异化学趋化性缺陷型，这类细胞能对一些刺激因子有反应，但对另一些刺激因子则无反应。

前面所描述的遗传实验技术和表现型分析结合起来，可以用来鉴定出涉及移动性和趋化性的基因。例如，通过分析特异趋化性突变体，可以鉴定出对特定刺激因子敏感的受体。

9.2.5 细胞分化

枯草芽孢杆菌的孢子形成

当一些细菌饥饿时，它们能够产生一个有抵抗力的内生孢子发生应答。在枯草芽孢杆菌（*Bacillus subtilis*）中，对孢子的形成进行了最广泛的研究，而其过程的调节已在第3章中进行了描述。在芽孢杆菌中生成孢子需要7h，以显微镜观察到的形态学变化为根据，这个过程可以被分成几个阶段（见图3.7）。不能形成孢子的枯草芽孢杆菌（*B. subtilis*）的突变株可以依照孢子形成被阻止的阶段来分类。因此，任何一个基因突变导致孢子形成在阶段Ⅱ受阻（不能前进到阶段Ⅲ）的突变体被称为*spo Ⅱ*，而那些孢子形成进行到阶段Ⅲ后才被停止的则称为*spo Ⅲ*突变体等。在每一个阶段，都有许多基因是必不可少的，如果这个过程要继续到下一个阶段，这可以通过互补和重组分析来辨别。这些不同的基因由字母和数字的进一步补充来表示。

例如，基于遗传分析，spoⅡ突变体的一种被命名为*spo ⅡA*。通过克隆一个可以和*spo ⅡA*突变体互补的DNA片段，可以分离到有关的基因。这个片段的序列显示它含有一个包括3个基因的操纵子，这3个基因分别被称为*spo ⅡAA*、*spo ⅡAB*和*spo ⅡAC*。*Spo ⅡAC*的序列和σ因子的很相似，已知σ因子决定RNA聚合酶和启动子的特异性结合。这提示*Spo ⅡAC*（现已知道是σ^F）在激活阶段Ⅲ所需要的前芽孢基因的表达上具有功能（见第3章），这随后也被纯化SpoⅡAC的体外实验所证实。另外两个*spo ⅡA*基因负责调节SpoⅡAC的活性，以便只在合适阶段SpoⅡAC被活化：SpoⅡAB是一个抑制SpoⅡAC（σ^F）的反-σ因子，SpoⅡAA是一个对SpoⅡAB起反作用的反-反-σ因子。

已经确定了大量孢子形成基因，带注解的基因组序列已经达到100多条。在某些情况下，已知某些基因的功能，可以用更具描述性的名称予以命名。例如，*spoⅡAC*被列为*sigF*。基因组序列的单独分析只能够对少量的基因功能进行临时的鉴定。而分离和描述大量的突变体则为理解孢子形成的过程打下了基础，特别是对这个过程的调节，以及第3章中描述的前芽孢和母细胞之间的交叉对答现象。但是，包括已知对孢子形成非常重要的很多基因的精确功能在内的很多方面的认识尚不清楚。

链霉菌的孢子形成

链霉菌(特别是天蓝色链霉菌)产生一种不同的孢子,这是一个散布机制而不是一个生存策略。这些细菌最初生长为丝状体,但是几天之后,它们分化为末端是孢子链的气生菌丝体。一些突变体(命名为 *bld* 或 bald)不能产生气生菌丝,然而另一类(*whi* 或 white,因为这个群体缺少和孢子联系的颜色)可以产生气生菌丝,却不能生成孢子。对这些突变体的分析表明,*bld* 基因包括一系列细胞外信号的分泌和应答所需要的功能,显示出一个气生菌丝的产生需要细胞间的信息交流,这和第 3 章叙述的群体感应系统相似。

在 *whi* 基因家族中,*whiG* 编码一个 σ 因子,在激活孢子形成的基因的表达转换中起作用,它与枯草芽孢杆菌和大肠杆菌中的另一种 σ 因子非常相似(见第 3 章)。基因组序列数据显示,在天蓝色链霉菌中控制孢子形成所需的许多基因,与在其他非孢子形成细菌中控制不同发育过程的基因有关。

通信和分化

一些细菌具有更复杂的生命周期,包括独特的形态学阶段(在某些情况下是多细胞行为)。这在简单的细菌系统和复杂的多细胞生物体的发育系统之间架起了一座有价值的桥梁。

一个例子是水生细菌新月柄杆菌(*Caulobacter crescentus*),它以两种形式存在:有杆细胞和游动细胞。前者依靠它的茎附着在培养基上(图 9.7)。随着它的生长,在茎的对面一个鞭毛在细胞的另一端形成,以便在细胞分裂时一个新的能动的游动细胞被释放出来,而其他细胞仍然附着在培养基上。这个游动细胞既不复制它的 DNA,也不进行细胞分裂,但是最终在一个新的地点定居下来,脱落鞭毛形成一个茎附着于新的培养基上。如

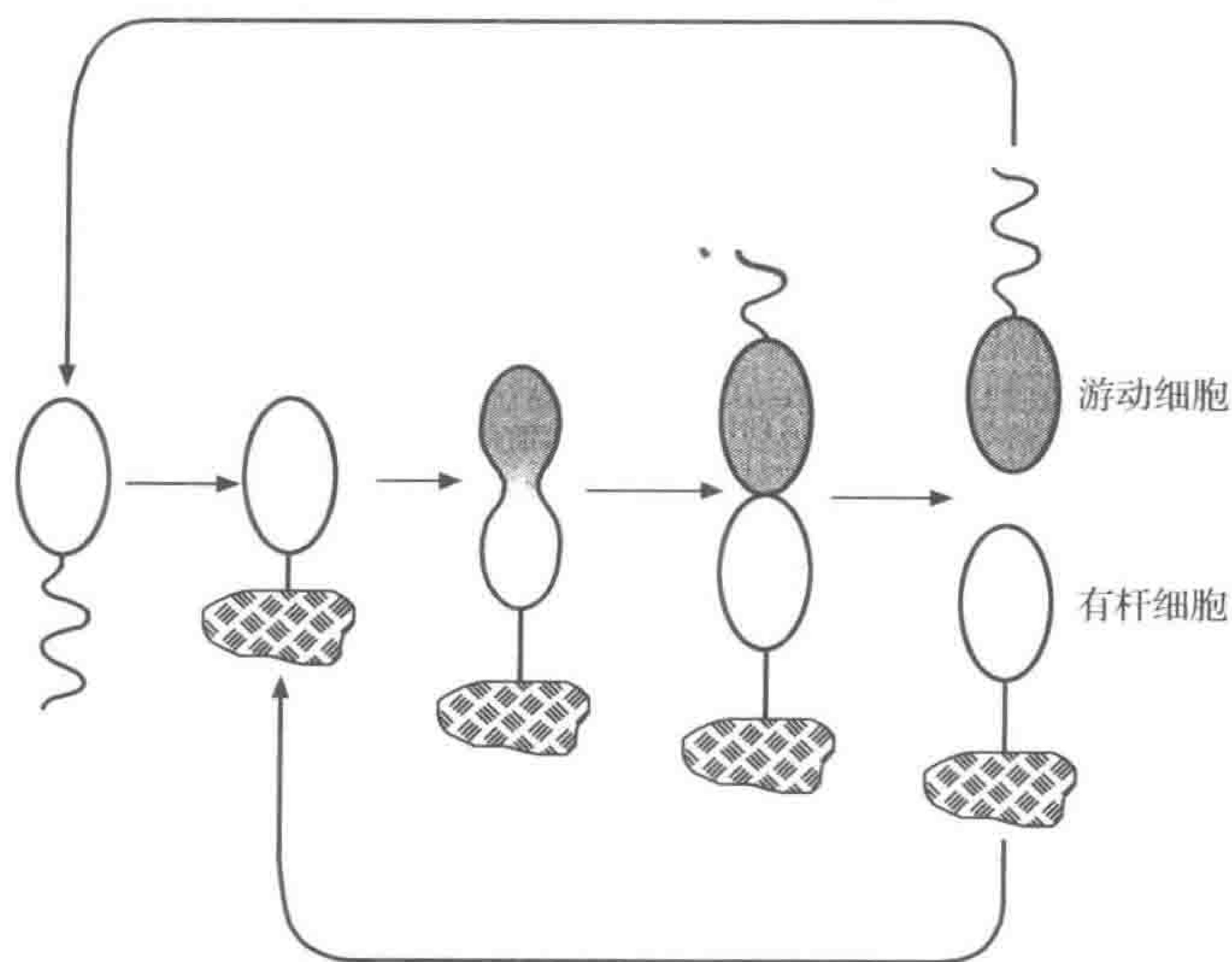


图 9.7　新月柄杆菌的分裂周期

有杆细胞,附着于表面,进行 DNA 复制和细胞分裂,产生一个能游动的、不能复制的游动细胞。最终脱掉鞭毛,附着于表面,并且变成一个可以复制的有杆细胞

今它已变成另一个有杆细胞，进行 DNA 复制并进入一个新的细胞分裂周期。关于控制这种行为的机制，可以提出一些有趣的问题，包括在两种类型的细胞中的基因表达调控和复制细胞发育不对称的原因。

报道基因技术和微阵列分析(见第 10 章)显示有杆细胞和游动细胞的基因表达具有差异。所涉及的调控基因之一是一个叫做 FlbD 的转录激活因子。激活状态的 FlbD 不均等地分布在分裂的细胞中，它只发现于游动细胞的极端，用于激活在游动细胞中产生鞭毛所需的基因。

另一个有着有趣生命周期的细菌是土壤细菌——黏细菌(*Myxococcus xanthus*)(同样见第 3 章)，其周期中包括多细胞行为部分。在有利条件下，单个细胞以协调的方式在表面群集。当食物变得稀少，单个细胞聚集形成隆起(包含大约 10^5 个细胞)，然后发展成含有大量孢子(黏孢子)的一个子实体(图 9.8)。黏孢子对干燥有抵抗力，并且能够使细菌存活直到条件改善，使得它们能够发芽并再次形成单个细胞。

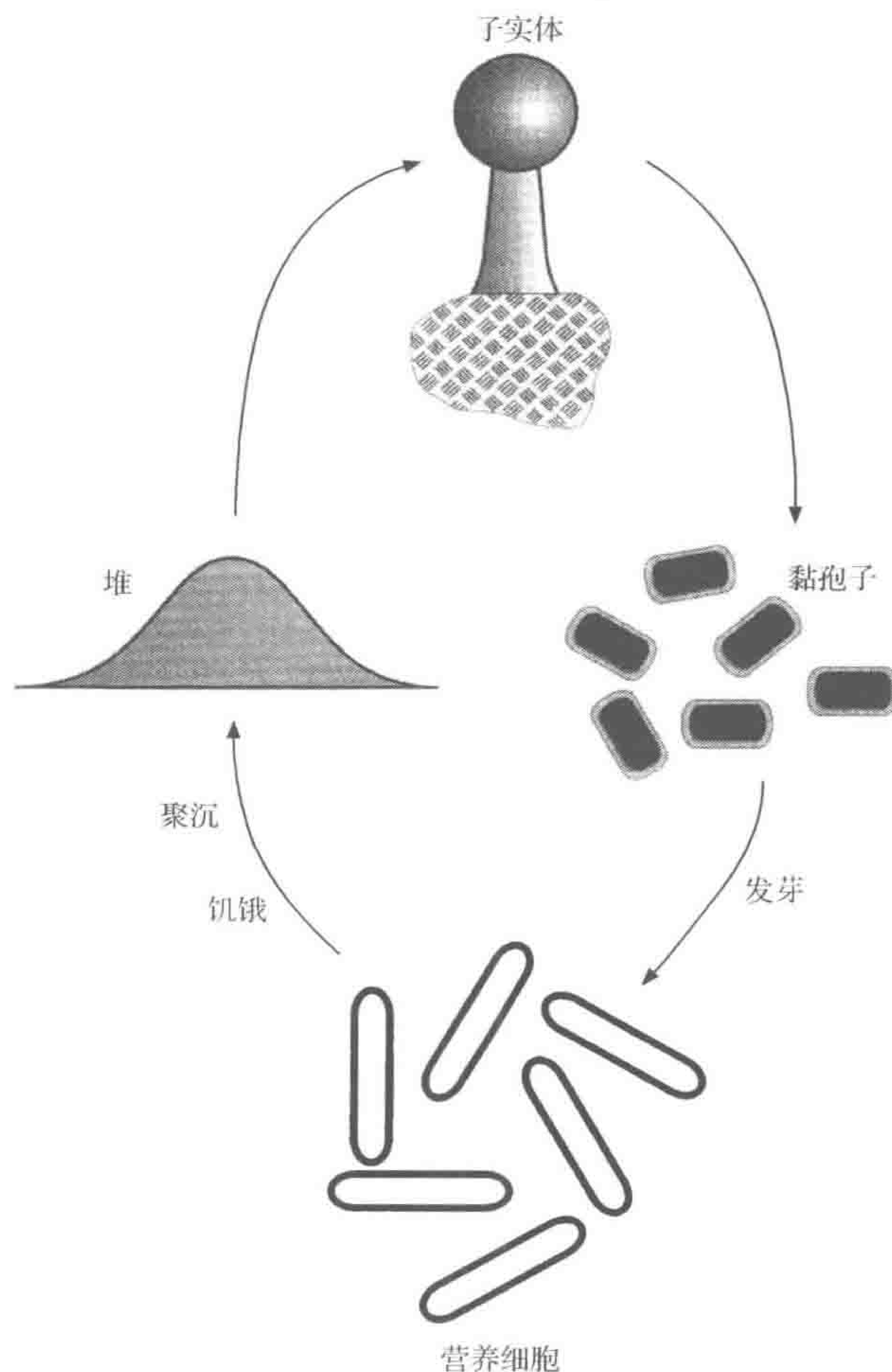


图 9.8 黏细菌生命周期的图解插图

这是一个广义和简化的图示。在不同物种中，子实体的结构各不相同，从简单到不包含杆结构到包含多个孢子囊的复杂的分支结构

从单个细胞群发展成一个聚集体，然后到孢子形成，是由一系列信号控制的。其中之一，A 信号是一种氨基酸的混合物，在这个过程的早期阶段，作为一个群体感应机制（如第 3 章中所述）起作用。聚合只在 A 信号达到阈值水平后发生，从而确保有足够的细胞形成一个有功能的子实体。进一步的发展是被另一个编码 C 信号的基因（*csgA*）控制的，因此，*csgA* 突变体的孢子形成也是有缺陷的。不同于被分泌的 A 信号，C 信号需要细胞接触，因为它是一种细胞表面蛋白，与另一个细胞的表面受体相互作用（图 9.9）。这种相互作用有两个主要后果。首先，它刺激 *csgA* 的转录，因此产生更多的信号，从而放大了效应。其次，它改变了 50 多个基因的调控，包括那些孢子形成所需的基因。

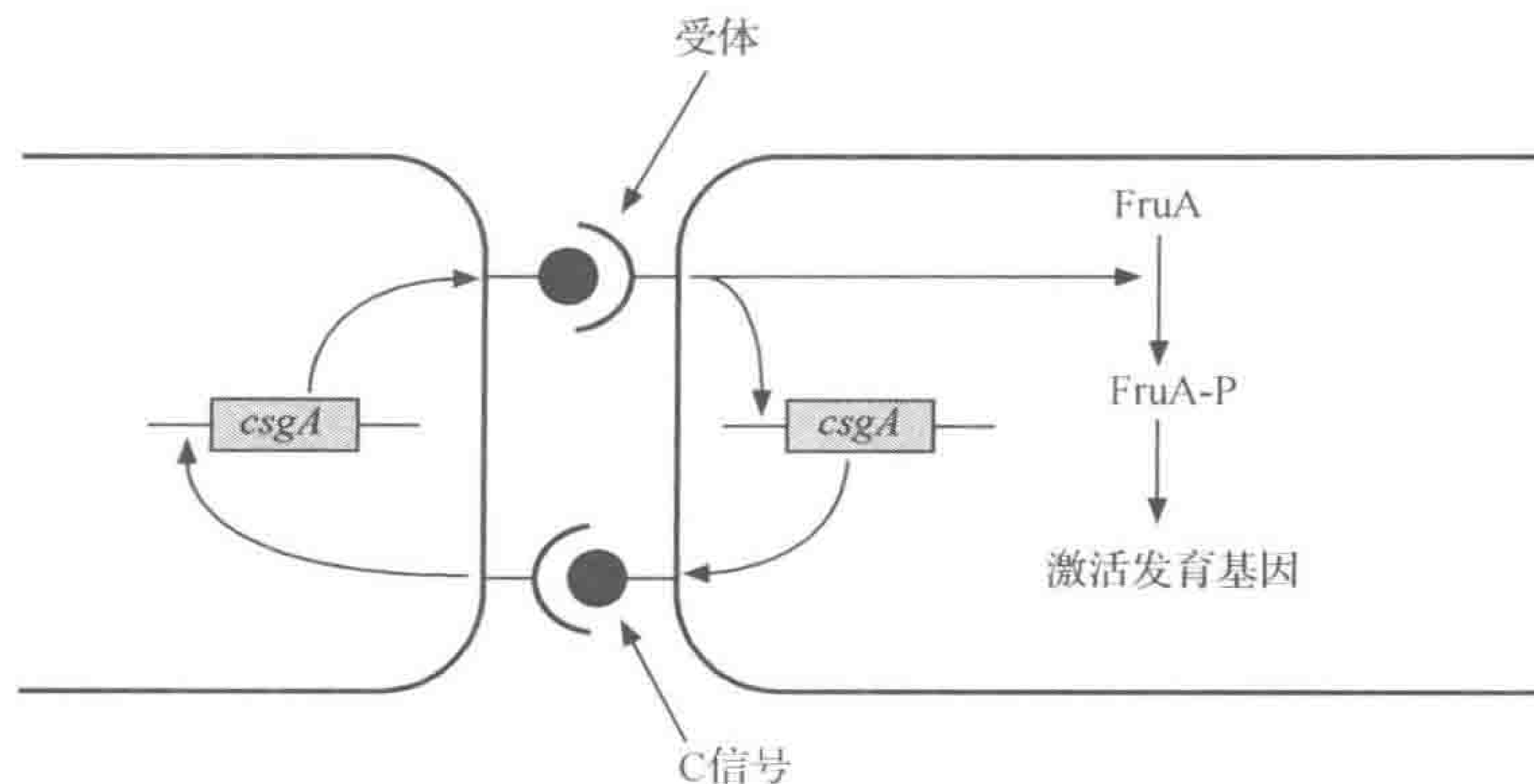


图 9.9 黏液球菌 C 信号通路的一些元件的简化模型

9.3 细菌毒力

9.3.1 细菌致病的全面机制

在感染过程中，细菌利用许多常见的分子机制达到不同的目的。本节利用细菌分子遗传学理论揭示微生物致病的关键机制。

从基因组序列信息可以很明显地发现，毒力基因经常会成簇集群出现，在与其密切相关的非致病性细菌中不存在这些区域。而且，基于这些区域和基因组其他部分的 GC 含量（见 9.5.2 部分）的比较结果，这些大的 DNA 片段（如 30～50kb）显然是从其他生物体中获得的。总之，根据这些观察结果提出了毒力岛的概念，表明细菌可以获得全部完整的系统以扩展其利用不同宿主环境的能力。第 4 章中提到的霍乱弧菌的例子，其毒力基因位于一个毒力岛上，随后证明它是一个整合的噬菌体。

当时没有想到，对植物病原体丁香假单胞菌相关的许多研究对致病性的理解具有显著的影响。在 20 世纪 80 年代中期，转座子诱变（见后面的章节）用于鉴别产生植物病害症状所必需的基因簇。利用穿梭载体将基因簇克隆到非致病性细菌中，发现这些无毒细胞获得了产生特定疾病症状的能力，因此，这个区域的重要性便得到证实。在 20 世纪 90 年代早期，逐渐发现，类似的基因簇存在于许多其他细菌病原体中，并且一些由这些基因簇编码的蛋白质装配成细菌鞭毛的输出工具的一部分。该信息形成了在细菌病原体中存在着常见的分泌途径的概念。

现在已经知道，这些区域编码一种称为Ⅲ型分泌系统的专门分泌装置(见第 1 章)。它们至少包括 20 种蛋白质，是细菌已知的最复杂的运输系统。它们通常被称为“分子注射器”，因为它们使革兰氏阴性菌分泌并将效应蛋白直接注入真核宿主细胞的细胞质。注入的蛋白质可以改变正常的细胞信号转导通路，导致宿主无免疫应答或导致建立细菌定植途径的细胞骨架重组。

大部分Ⅲ型分泌装置在亲缘关系较远的相关病原体之间是保守的(虽然效应蛋白不同)。因此，从人类淋巴腺鼠疫到番茄植株南部枯萎等多种疾病，具有和普通细菌相同的致病机制。例如，耶尔森菌，注入至少 3 种不同的效应分子才能破坏免疫细胞的重要功能。编码Ⅲ型分泌系统的基因，包括效应蛋白和结构蛋白的基因，都聚集在一起，可作为细菌种间转移毒力岛的另一个例子。

9.3.2 毒力基因的发现

许多细菌病原体有独立的自由生活和致病的生命周期，所以它们遇到不同的环境，需要发挥不同的功能以生存。因此，病原体必须能够在宿主内识别表达毒力基因所需要的信号。例如，志贺菌属(*Shigella*)毒力基因功能的表达是由体温(37℃)所引起，而非环境温度(<30℃)。因此，如果能识别仅在感染过程中表达的基因，那么也可以识别其重要的毒力特性。报道基因特别适用于这一目的。

体内表达技术

应用报道基因识别致病基因最广泛使用的方法之一就是所谓的体内表达技术(IVET)。该技术选择只能在宿主体内表达的细菌启动子，并由此驱动毒力特征的表达。在 IVET 的大多数例子中，来自细菌宿主的随机 DNA 片段插入相邻的无启动子的报道基因中，那些基因产物具有一定的表型，可以在宿主体内进行阳性选择。例如，由于嘌呤的生物合成对环境中沙门菌至关重要，*purA* 突变的鼠伤寒沙门菌(*Salmonella typhimurium*)不能在动物模型中生存。通过携带功能性的 *purA* 基因质粒可以补充这种缺陷，只有在体内插入具有功能的启动子时，该质粒才能表达有功能的 *purA* 基因(图 9.10)(构件中包含 *lacZ* 基因，其表达受插入的启动子控制，因此，可以通过 *lacZ* 基因的表达，鉴定出组成型的启动子，并将其排除在外)。当用携带不同 DNA 片段的克隆库感染动物时，只有那些含有活性的启动子克隆能够生存，并且能从动物中回收，进而可以直接选择在感染过程中具有活性的启动子。对这些启动子片段的鉴定导致对其正常调节基因的鉴定，其中可能包括沙门菌毒力所必需的基因。

转座子突变

正如前面提到的，鉴定负责某一特定表型的基因的典型方法是产生有缺陷的突变体。对于毒力基因，突变体的表型在疾病的体内模型中不能生存，如动物模型或组织培养细胞。常以不同形式利用转座子进行诱变。

因为可以通过插入使基因失去活性，也可以很容易地鉴别没有活性的基因，所以，转座子提供了一种产生突变体的方便手段(见第 7 章)。图 9.11 中所列出的一个典型策略

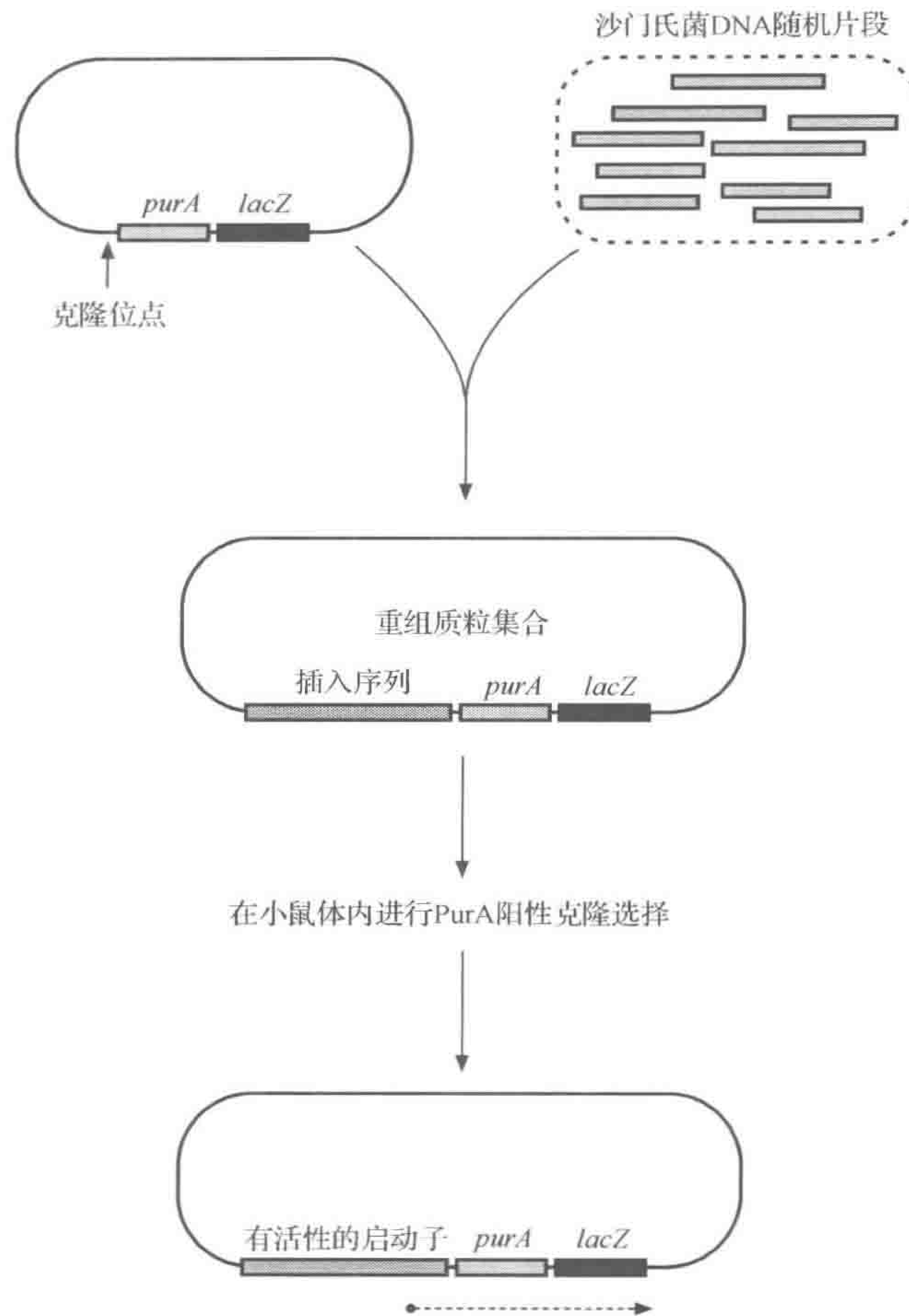


图 9.10 体内表达技术(IVET)

不含启动作用的 *purA* 和 *lacZ* 基因的质粒载体，这些基因的表达依赖于具有启动子活性的 DNA 片段的插入。用包含重组质粒的沙门氏菌感染的小鼠用于在这种条件下具有活性的启动子 $PurA^{+}$ 克隆的选择

就是应用温度敏感的载体构建携带转座子的质粒，这种质粒在温度升高时不能复制。当将质粒引入细菌细胞时，在质粒可复制的温度(30℃)条件下，细胞能够生长且所有的细胞都携带质粒。如果随后将培养温度改变到质粒不能复制的 42℃条件时，只有那些转座子成分已转移到染色体上，保留转座子的细胞可以生长。如果应用像 Tn*5* 这样的携带有卡那霉素抗性的转座子，则可以在 42℃条件下，应用含有卡那霉素的琼脂平板培养细菌，筛选整合了转座子的细胞。

将转座子插入染色体的不同位置，可以产生转座子突变体文库。利用转座子所携带的抗性基因选择所需的突变体，然后可以确定转座子插入的基因。如果构建一个基因文库，并涂布在含卡那霉素抗性的琼脂板上培养，仅那些包含 Tn*5* 的克隆能生长。这些克隆同时也包含被转座子插入的基因，它们是在 Tn*5* 插入后失去活性的，故可以描述其特性。

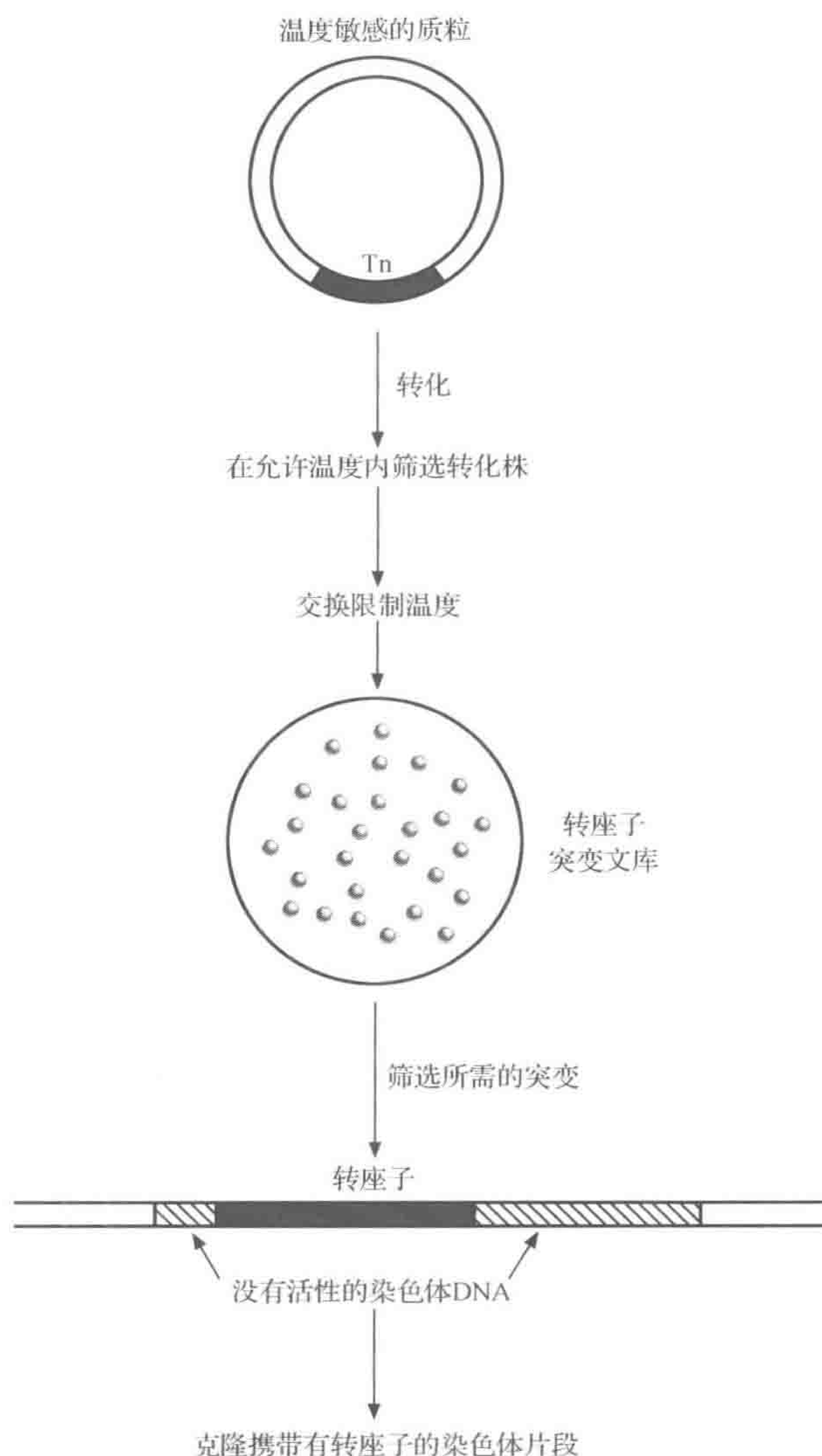


图 9.11 转座子诱变

一些转座子随机插入到染色体 DNA 上的能力用于产生突变体文库。
转座子诱变的一个优点是很容易鉴定受影响的基因

另一种鉴别转座子插入位点的方法是应用 PCR 突变体(见第 2 章),称为反向 PCR。常规 PCR 的局限性是仅扩增已知引物序列之间的 DNA 片段,很难分析未知的序列。在插入转座子的情况下,该转座子的序列是已知的,可以作为靶向序列的 PCR 引物,但是由于转座子插入后其侧翼序列是未知的,常规 PCR 反应所需的第二对引物不能设计。反向 PCR 是用于扩增一段已知序列侧翼的未知序列(图 9.12)。对于转座子突变体,部分限制酶切染色体 DNA 的片段中包括完整的转座子靶序列,用 DNA 连接酶连接呈环状。随着 DNA 序列呈环状,未知序列连接到已知的序列中。应用转座子特异性的引物,以这些分子作为 PCR 反应的模板,直接进行 DNA 的合成。因此,未知序列得到扩增。扩增片段

被克隆和测序，以确定转座子的插入位点。

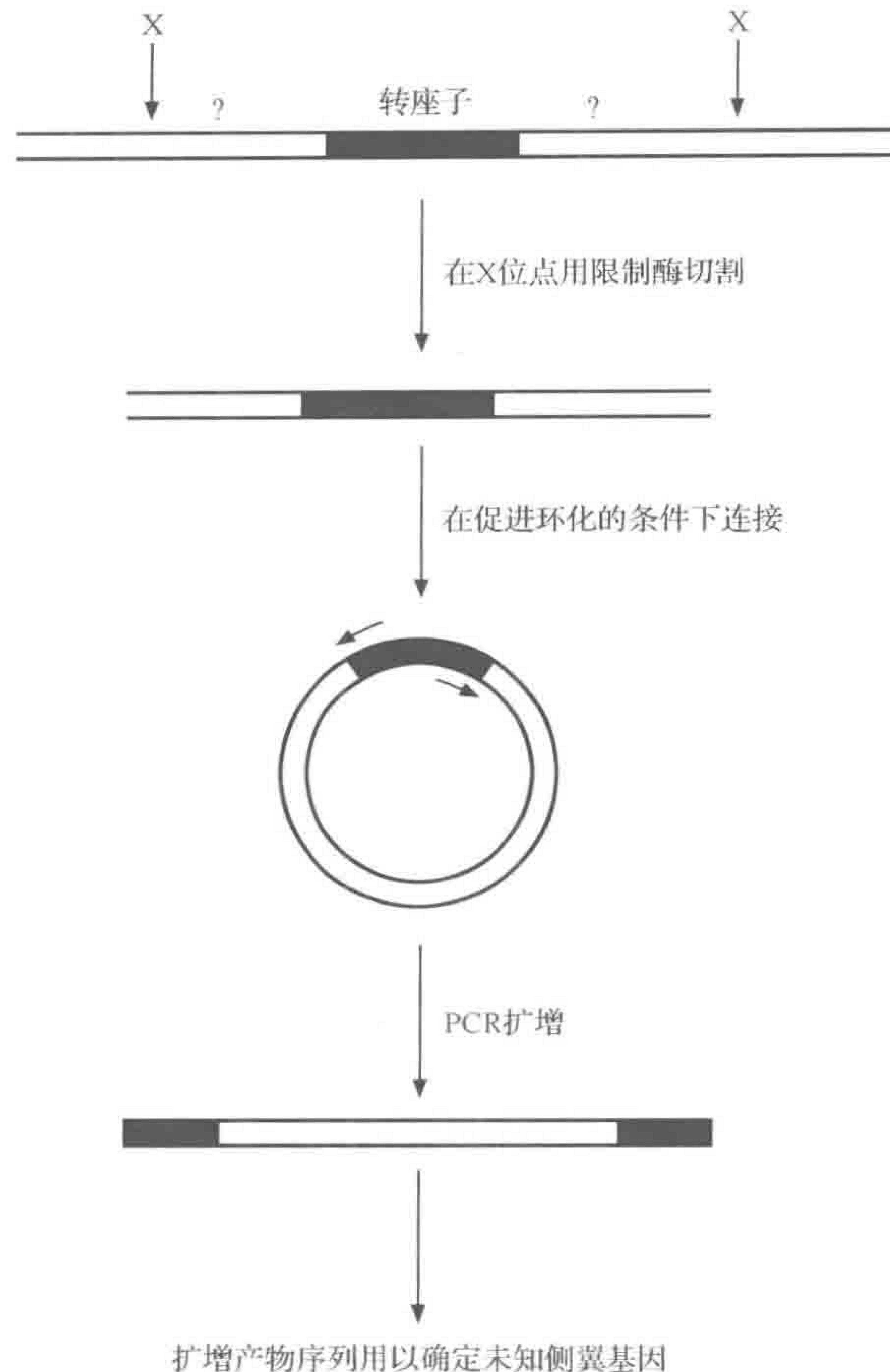


图 9.12 反向 PCR 用于确定插入的转座子侧面的未知 DNA 的应用
限制性片段的末端的连接意味着转座子特异性引物可用于扩增侧面的序列

信号标签诱变技术

转座子突变是诱导和鉴定不同表型突变体的有力手段。然而，由于在感染的动物模型中分离的突变体细胞恰好不能生长，它在细菌毒力的研究的直接应用中是有局限的。为了避开这一障碍，科学家发明了信号标签诱变技术（STM）。它来源于转座子突变（见上文），是一种重要的负性选择技术，用独特的 DNA 信号标记其不同的单突变体。通过对来自那些康复的感染模型的最初接种物的突变体进行比较，可以识别那些不能幸存的突变体，从而鉴定毒力所需的基因。

图 9.13 给出了 STM 的概况。构建转座子突变体文库，每一个突变基因被一段 DNA 标记，该 DNA 片段和其他标签具有共有序列，均包含一个独特的中心区域和两个侧臂。STM 的关键是，每一突变个体拥有自己独特的标签，可区别于其他突变体。突变单体单独保存，并且每一突变体 DNA 以网格状的方式点到膜上。突变体混合并接种相

关动物模型，回收能够生存并感染动物的细菌。利用 PCR 扩增回收细菌中的所有标记，其混合产物作为探针与网格膜杂交。因产物池中不含无法在感染动物体内生存(那些毒力缺陷)的突变体的标签，所以它们将被识别。可以从原始的储备阵列中回收它们，并进行深入的研究。该系统已广泛应用于细菌定植、免疫系统逃避和人体细胞附着相关的新毒力因子的鉴定。

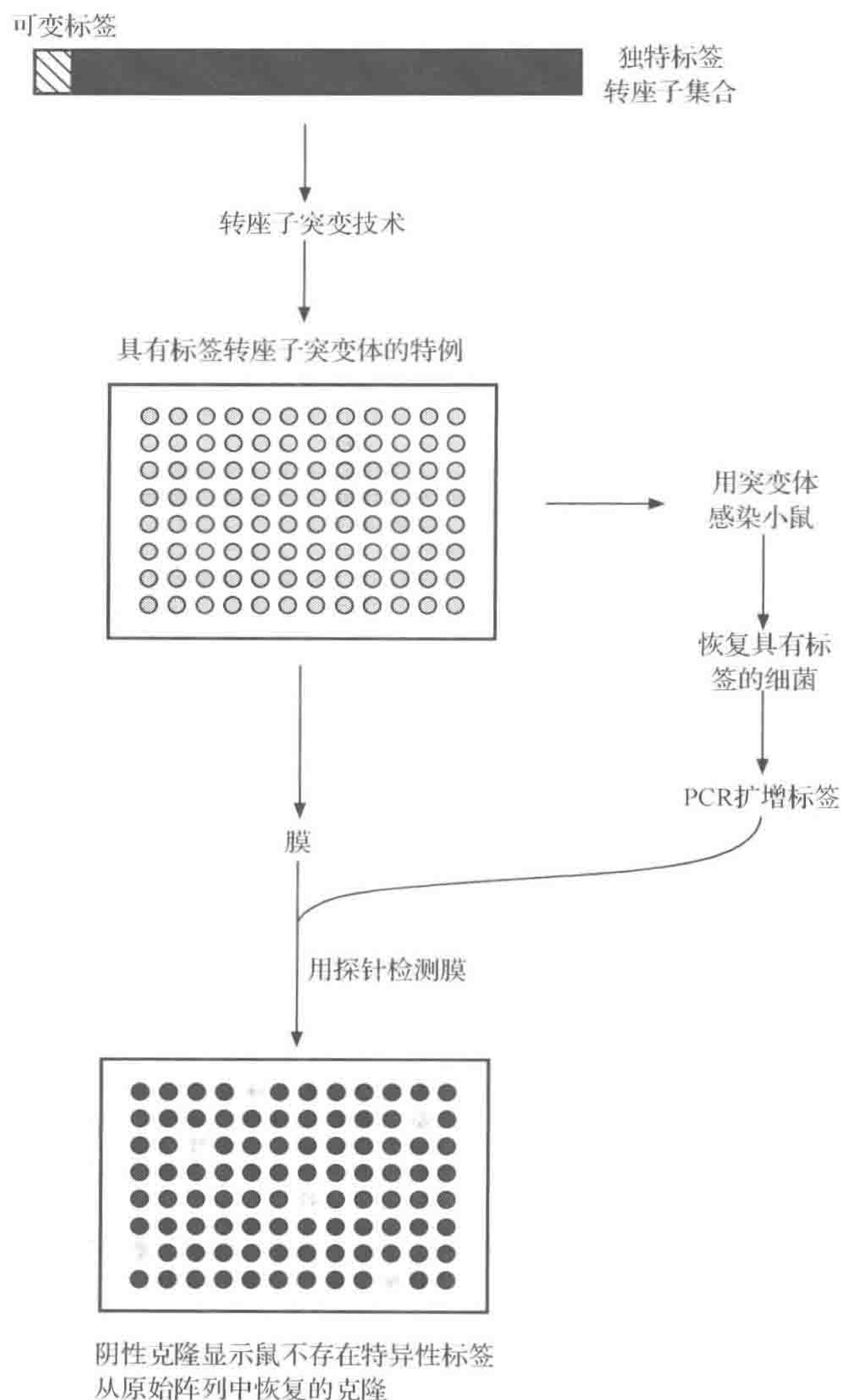


图 9.13 信号标签诱变技术

使用转座子突变使感染小鼠所需的基因失去活性，并提供了鉴定那些基因的一种方法

转座子原位杂交

基于转座子的另一种技术是转座子原位杂交(TraSH)，该技术可常规识别在一定的选择条件下不能生长或生存的突变体，如图 9.14 所示。本系统利用一种专门设计的转座

子,其含有转录方向向外的 RNA 聚合酶启动子,从转座子文库的突变体中分离 DNA 后,转座子插入位点的侧翼基因可转录形成与之互补的 RNA。

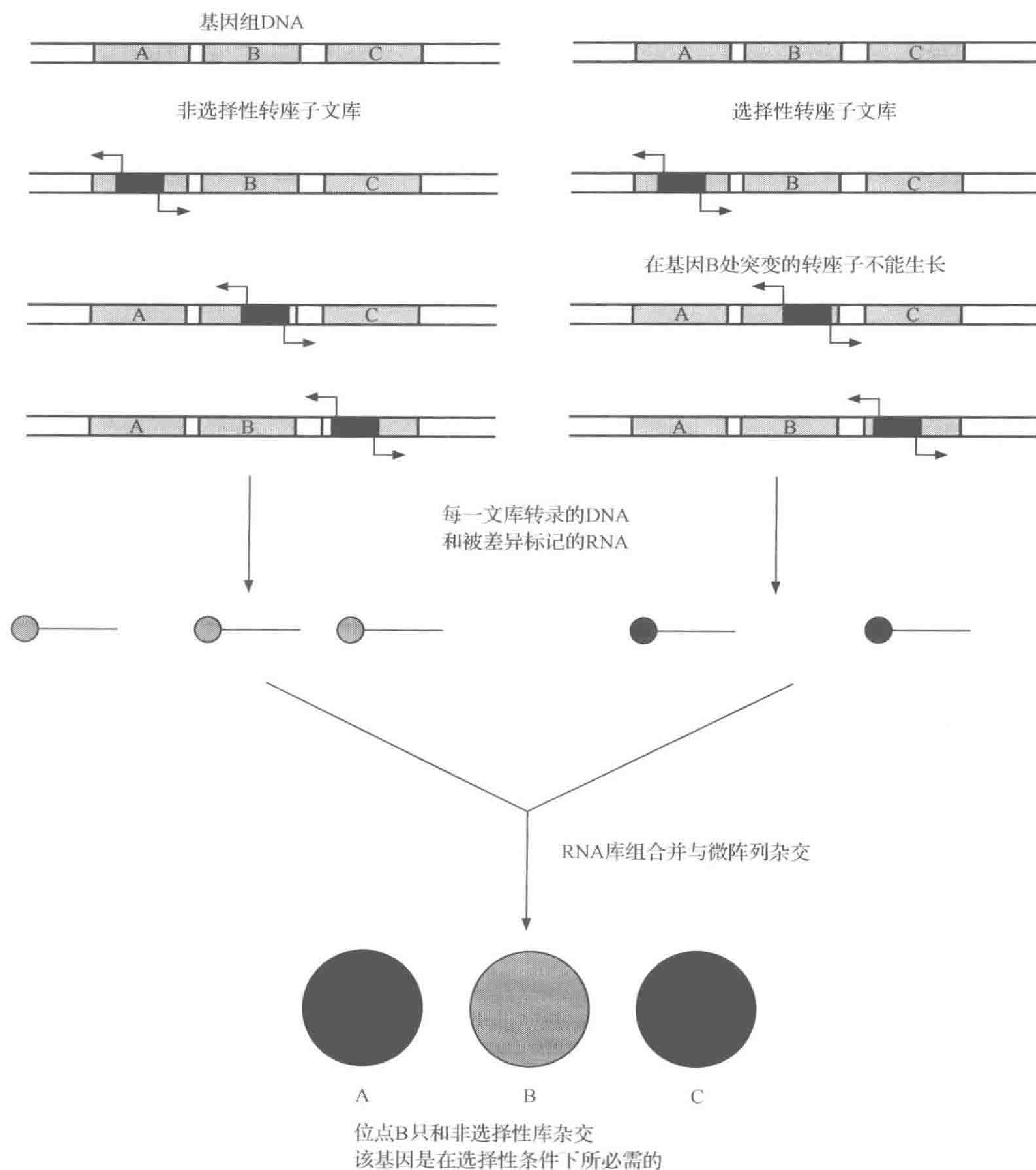


图 9.14 转座子原位杂交的原理

转座子携带用于侧面序列转录的向外的启动子。来自选择性和非选择性池的不同标签的转录体能够鉴别不能在选择性条件下生长的突变体,那些杂交点将不能从微阵列中显示出来

与 STM 类似,构建转座子突变体文库并分成两组,一组在标准条件下生长,另一组在一些选择性条件下生长(如可能在动物模型或组织培养细胞中生长),分离获得两组 DNA,利用转座子上的转录方向向外的 RNA 聚合酶启动子,将每一插入片段两侧的 DNA 转录成 RNA。在体外转录过程中,用不同的荧光标记不同的 RNA 混合物,因此,

RNA 将带有不同的标签，在代表实验细菌的整个基因组的两组 RNA 微阵列表面结合并洗脱(见第 10 章和前面提到的 ChIP-on-Chip)。

在正常或选择性条件下同样生长的转座子突变体，其细胞数量及其两组特殊标记 RNA 的浓度大致相同，结合到相应微阵列点上的两种不同标记的 RNA(两组)的数量相当。然而，在选择性条件下不能生长的转座子突变体，在混合物中将不存在该插入位点侧翼基因的 RNA，因此，只有来自正常生长池标记的 RNA 能结合微阵列的同源靶序列。

微阵列结果可以用于明确地鉴定在一种条件(而非另一种条件)下生长所需要的基因。在实践中，该技术已被用于鉴定致病性细菌在巨噬细胞内生长所需的基因，巨噬细胞作为选择性条件，而实验室生长培养基作为标准生长条件。

9.4 特异性突变

到目前为止，本章介绍了随机突变的产生，以及在改变表型的基础上进行选择的突变技术，可鉴定相关基因并且确定基因的功能。从一个特定基因参与的假定开始，利用 DNA 重组技术可以完善这些技术，通过修饰或删除该基因，鉴定其表型变化。因此，常规的基因组学从表型开始，其后续工作是鉴定与基因有关的特性和功能，而分子方法是从改变基因开始，然后分析其表型。

9.4.1 基因替代

鉴定特定基因功能的关键技术是通过各种已知的方法，包括基因替代，同位基因替代或基因敲除的方法，使目标基因失去活性。本质上都是应用同源重组移除全部或部分特定基因，或用改造的、或没有活性的基因取代特定基因。图 9.15 介绍了该过程是如何完成的。在这一例子中，构建克隆基因中间部分被除去并被抗生素耐受基因(*aph*，氨基糖苷磷酸转移酶，具有卡那霉素抗性)所取代的质粒。卡那霉素抗性基因侧翼序列与宿主菌株相同，同源重组在此区域内发生。注意两次重组(双重交换)是基因替代所必需的。单个位点的重组(单一交换)只是将质粒整合到染色体上。需要另外的技术确保获得双重交换。

在选择的宿主细胞内，所使用的质粒不能复制。转化以后，利用卡那霉素抗性可以筛到通过靶基因同源重组已将 *aph* 基因整合到染色体上的细菌，因此，*aph* 基因的插入使靶基因失去活性。接下来可以检测是否获得预期的表型。例如，如果认为靶基因是菌株在巨噬细胞内生存和生长所必需的，可以检测突变体是否在这方面具有缺陷。

结果并不一定绝对是事实。基因以许多方式相互作用，敲除一个基因会对其他基因有间接影响。如果靶基因是操纵子的部分基因，在操纵子内部，基因敲除后尤其会影响其他基因的表达(换句话说，突变是直接对立的)。可以通过互补，部分消除这种可能性。构建一个携带野生型基因的质粒(在宿主内能复制的质粒)，然后将质粒引到突变体中。如果原来的基因敲除后导致可观察到的效应(如不能在巨噬细胞中生长)，则引入质粒后将恢复最初的野生型表型。成功的互补支持这一概念，该基因产物是菌株在巨噬细胞中生存所必需的。但是需要指出的是，它仍然不能最终确定该基因产物的直接作用，例如，它

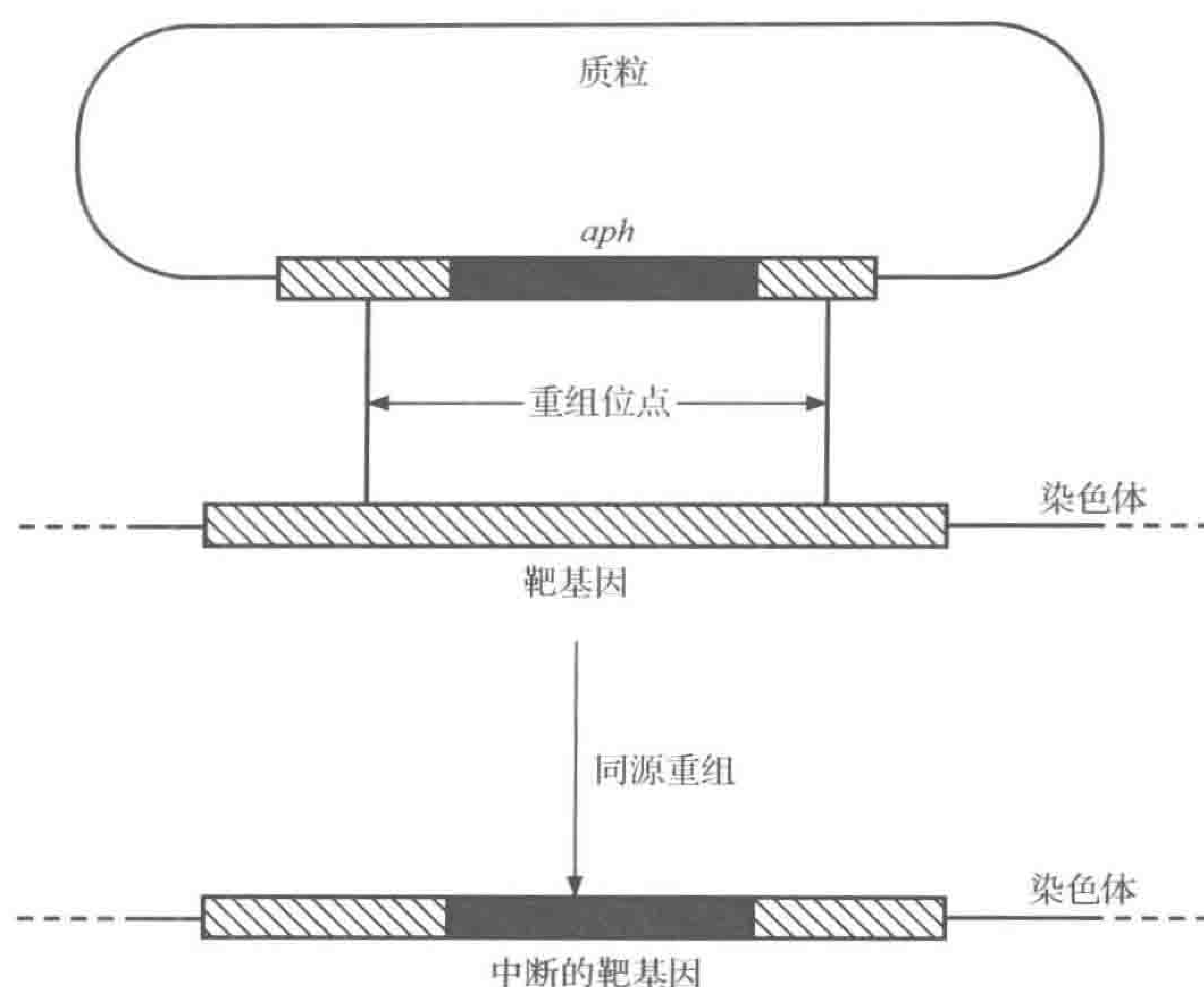


图 9.15 等位基因替代的基因破坏

在破坏的克隆基因和染色体上靶基因的同源重组导致靶基因失去活性。
aph，氨基糖苷磷酸转移酶（卡那霉素抗性）

可以通过改变其他基因的表达来发挥功能。

在第 10 章中，描述了该技术是怎样用于替代特定突变的基因。

9.4.2 反义 RNA

基因敲除的一个问题是如果在实验室条件下，某一基因是细胞生长所必需的，完全去掉该基因，细胞将是致死性的，以至于不能获得任何重组子。因此，减少基因的表达或产物的功能是有用的，这样可以留下足够的活性在适宜的正常实验室生长条件下生长，但是不足以应对随后强加的压力条件。一个办法就是使用反义 RNA。

为了达到这一目的，将基因的一部分克隆到表达载体，它由载体上的启动子驱动转录，但是应按错误的方向插入。因此，插入片段按与正常方向相反的方向转录；换句话说，相反的链得到转录。产生的 RNA 链（反义 RNA）和正常的信使 RNA 互补，二者配对将产生双股 RNA 分子，进而干扰信使 RNA 的翻译，故蛋白质表达水平下降。使用强的或弱的启动子（或甚至更好的能打开和关闭的启动子）将导致不同数量的反义 RNA 的产生，因此，将改变所形成蛋白质产物的减少数量。

在第 3 章中，介绍过 RNA 相关的活性调节。在第 5 章中，介绍了质粒复制控制相关的 RNA。

9.5 分类学、进化和流行病学

9.5.1 分子分类

细菌传统分类的依据是按照生物多样性背景条件下将具有相同特性生物体进行分组的原则。应用这一方法，很可能是使用一系列复杂的生化实验区分两种相近的相关生物体。然而，这一方法并不一定提供不同生物体的精确进化关系。传统分类学因此会导致一些不大相同的生物体由于相同的种或属被错误地分到一组。分子途径在解决上述问题中发挥了重要作用，下面将做具体讨论。

9.5.2 GC 含量

用于细菌分类的一个简单的分子特性就是 DNA 的碱基组成，规定为鸟嘌呤和胞嘧啶残基的数量组成占碱基总数的百分比(GC%)。尽管基因组测序能给出最准确和精确的值，但不必通过基因组测序也可以测定它的值，即使用物理方法确定 GC 含量。细菌 DNA 的碱基组成差异很大，从一个物种到另一个物种跨度为 20%～80%，但密切相关的生物往往有相似的 DNA 碱基组成。对此有很好的解释，复制、转录和翻译在不同程度上以不同的方式对核酸的碱基组成很敏感，所以可以一起进化。

如果利用这一概念进行细菌分类，有时会发现，在其他方面相似的生物体却有相当不同的 DNA 碱基组成。例如，金黄色葡萄球菌属和微球菌属在形态上与革兰氏阳性球菌相似(虽然可以用生化方法区别它们)。但是微球菌 GC 含量(约 70%)较高，而金黄色葡萄球菌 DNA 的 GC 含量的比例低(30%～39%)。

9.5.3 16S rRNA

核糖体 RNA 测序是一种更强大的技术。核糖体 RNA 基因是高度保守的，在所有细菌中显著相似，但从一个物种到另一个物种，核糖体 RNA 基因的序列也有小的变化。使用这些变化(最常用的 16S rRNA)，不仅可以区分不同的物种，而且可以发现差异的程度。换句话说，通过计算两个物种之间碱基的数目不同，可以获得将它们分开的进化距离。如果对相当数量的序列进行比较，可以构建系统进化树，它会显示这些物种从一个共同的祖先进行分化的可能路线。一个简单的例子如图 9.16 所示(用比在实践中使用的更短的序列)。与 C 或 D 相比，生物体 A 的序列更相似于 B，这在进化树的排列上可以反映出来。值得注意的是：构建进化树的复杂程度远远超过简单的描述，从一系列数据中可以构建得到许多进化树。计算机选出最适合的数据，但它只是一个模型，并不一定反映生物体所涉及的真实的进化。

分析细菌 16S rRNA 基因是一种有效和广泛应用于建立不同细菌间的亲缘关系进化的方法。但 16S rRNA 基因的进化树是否真的反映了生物体的进化史呢？许多细菌横向基因转移(毒力岛、温和性噬菌体、质粒等)的例子足以表明，没有建立在单个共有基因序列比较基础上的进化树能可靠地反映细菌进化的复杂性，也包括 16S rRNA 基因。为了

微生物	序列
A	C A U A G A C C U G A C G C C
B	C C U A G A G C U G G C G C G
C	C G A A A G G G U G G C G C C
D	G C U A G A A G U G G C G G C

举例中仅是短的序列，实际上将使用更长的序列

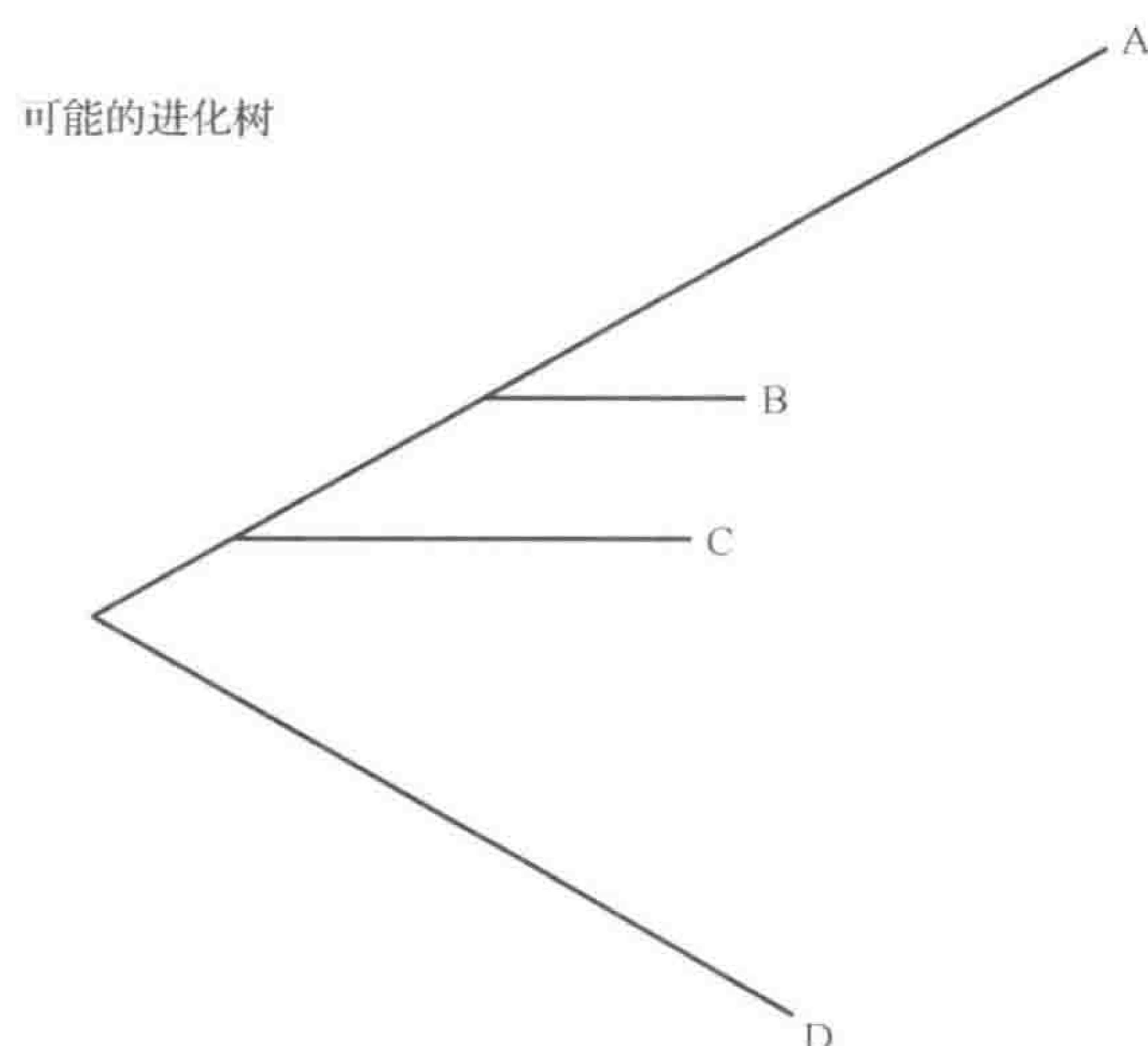


图 9.16 序列数据库的系统发育进化树的构建

这仅仅说明了原理。实际上，为了检测所画的许多可能的进化树，将需要使用更长的序列，进行更广泛的计算机分析

描述真实的进化发生，包含整个基因组信息的方法正在得到发展，将来可能比较不同基因组之间许多共有的基因，以便建立更精确的进化。

当进化树仅以 16S rRNA 序列为基础时，尽管可能存在问题，但鉴定细菌时不必克隆 rRNA 基因，故其依然是强有力的工具。16S(或 23S) rRNA 基因的突变并不能均等地出现在整个基因中，一些区域是相当保守的。因此，可以用一对能识别位于突变区域两侧保守序列的 PCR 引物扩增包含不同基因的区域，然后对扩增产物进行测序。保守程度如此之高以至于使用相同的一对引物可以对未知的任何生物体进行扩增。然后通过与已知生物体的 rRNA 序列的比较，至少可以暂时地测定未知细菌的特性及其与已知物种的关系。

PCR 能够扩增微量 DNA 的强大功能意味着不再存在培养细菌的问题。这一点很重要，因为标准细菌学技术需要培养一定量的细菌，尤其是医学上重要的病原体。虽然可以使用不同的培养基和不同的生长条件扩大分离细菌的范围，然而，即使拓宽使用条件的范围，仍然会有绝大部分细菌无法生长。设计针对 16S rRNA 基因的引物，对该样本进行 PCR 扩增，将会产生很宽范围的扩增产物。像构建基因库那样(第 8 章)，克隆这些混合产物，分离每个基因并测序，因此，可以鉴定存在于样品中的细菌(在数据库中的已知序列

的有限范围内)。例如,利用培养技术对人类结肠菌群进行了广泛的研究,并且它的组成细菌的特点被认为已经得到了全面的描述。然而,上文所述的基因型方法表明,76%的16S rRNA基因并不能与已知的生物体合适,这显然是来源于迄今未知的和不可培养的细菌种类。应用该技术和其他以序列为基础鉴定环境样本中细菌的方法,包括那些不能培养(宏基因组学)以及与其相关的一些技术,将在第10章中详细介绍。

9.5.4 变性梯度凝胶电泳和温度梯度凝胶电泳

对应用上述PCR方法产生的16S rRNA基因片段进行测序,可以明确地给出最终的鉴定结果,但这个过程相当复杂。要通过这些基因片段给出复杂细菌集落组成的重要信息注定是昂贵而且耗时的。设计靶向16S rRNA序列的引物,PCR扩增环境DNA样本,将产生代表许多在同一样本中占主导的微生物体的DNA序列的混合物。这些片段可能包括不同的序列,以帮助鉴定产生DNA的细菌,因为16S rRNA的保守性较高,通过常规的琼脂糖凝胶电泳不能够分离几乎具有相似大小的PCR产物。如果在普通的琼脂糖凝胶上分析,将得到单一的无信息的条带。

以序列不同为基础分离PCR产物的变性梯度凝胶电泳(DGGE)和温度梯度凝胶电泳(TGGE)两种相关技术,根据DNA变性特性的不同,可以解决普通琼脂糖凝胶无法分辨相似大小分子的局限性。在两种方法中,在聚丙烯酰胺凝胶电泳时,迁移的DNA片段处于变性梯度中。在TGGE中,变性梯度将逐步升高温度,或在DGGE时,会出现化学变性梯度。当DNA片段达到特殊的变性点时,双股PCR产物较弱的解链区域开始变性。它在胶内的迁移明显变慢,其他不同的PCR产物将继续。在不同条件下,来自不同细菌的变性DNA最终形成许多不同的条带,每一条带代表着存在于最初的集落里的不同细菌株。因此,对存在于任何选定的环境下的主要微生物集落成员的丰度和富集程度,这两种技术可以提供粗略的估计。

9.5.5 应用PCR进行诊断

用于细菌检测和鉴定的传统方法依赖于对生物体的纯培养生长状况,并结合染色方法、生化反应和其他检测方法。这些方法同样适用于环境生物检测(土壤或水)、食物中的细菌(包括牛奶和饮用水)或传染病患者样品中的病原体。然而,这些方法速度慢,一般至少需要24h,对于生长缓慢的生物,如结核分枝杆菌(*M. tuberculosis*),则需要几个星期。此外,还有一些细菌,如麻风(*Mycobacterium leprae*)(麻风病的病原体)是不能在实验室条件下培养的。

原则上,应用基因探针直接检测样本中特定的DNA可提供更快的结果。然而,该方法仅适用于大量细菌存在时。对于检测存在于这些样本中的重要但少量存在的生物体,基因探针则不够灵敏。所需要的检测技术是在第2章中所述的聚合酶链反应(PCR)。在理论上,它能检测出一个生物体的存在,这大大提高了灵敏度。为了将PCR应用于一种特定菌种的检测,需要知道该菌种典型的基因序列,也就是说,它是该菌种始终存在的(序列是保守的),但在其他细菌中则不存在或与其他细菌显著不同。然后设计结合到靶序列

上的一对PCR引物，可很容易地检测PCR扩增所产生的DNA片段。其他缺乏这些引物的特异性结合位点的细菌则不会产生扩增产物。

在研究实验室中，主要通过凝胶电泳检测扩增产物，有时也与Southern印迹结合，通过与特定的基因探针杂交(见第2章)，可以增强操作方法的敏感性和特异性。一些检测细菌病原体的商业化试剂盒，采用不同于PCR的基因扩增形式，而是利用其他更快的方式检测扩增产物。在第10章中，将介绍已知的实时PCR，它是比凝胶电泳更快速地获得结果的一种技术，其突出的优点是能定量，也就是说，它能直接告知样本中目标基因的量。

9.5.6　分子流行病学

流行病学是研究疾病的发生和分布的学科。通过查明传染源，可确定必要的措施以控制疫情暴发。微生物实验室通过鉴定病原体并确定相关的菌株(菌株类型)达到这个目的。同一来源的疾病患者将感染相同的菌株，如果菌株不同，病例将不属于同一疾病暴发。

有许多方法可用于不同的细菌种的分型。在第4章中介绍了利用噬菌体将金黄色葡萄球菌分型。使用可变抗原的血清分型也被广泛应用。近年来，分子分型方法已经越来越受欢迎，它能可靠地鉴定相同菌种的不同菌株之间DNA序列的微小差异。

限制性片段长度多态性

限制性片段长度多态性(RFLP)是最广泛使用的方法之一。当基因探针与来自不同菌株的限制酶消化的DNA(见第2章)进行Southern印迹时，不同菌株的检测片段大小可能会有所不同。这种效应可能是由探针检测区域内的DNA片段产生点突变、改变(或产生)限制性位点、插入或缺失所导致的(图9.17)。

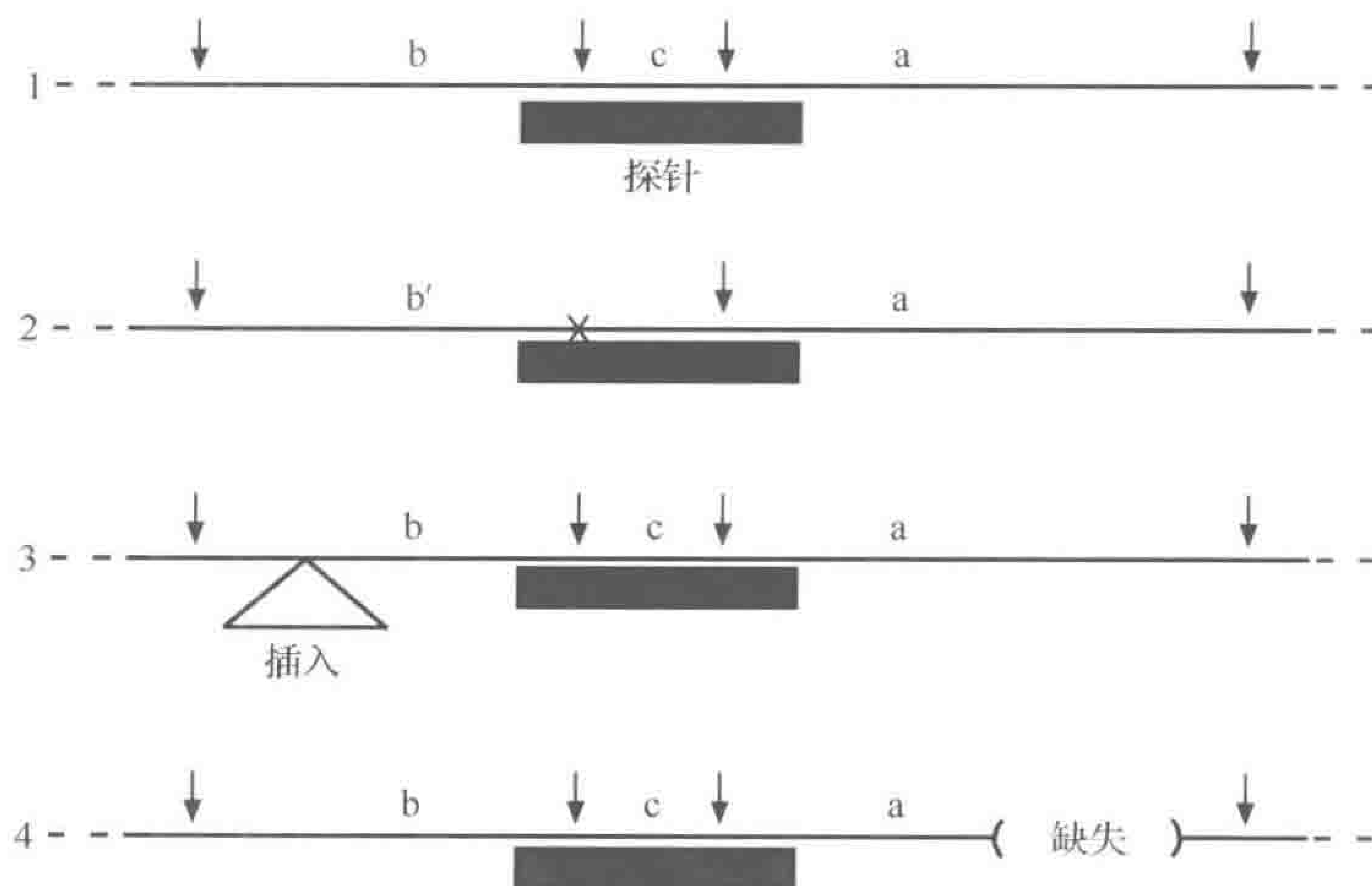


图9.17　限制性片段长度多态性(RFLP)

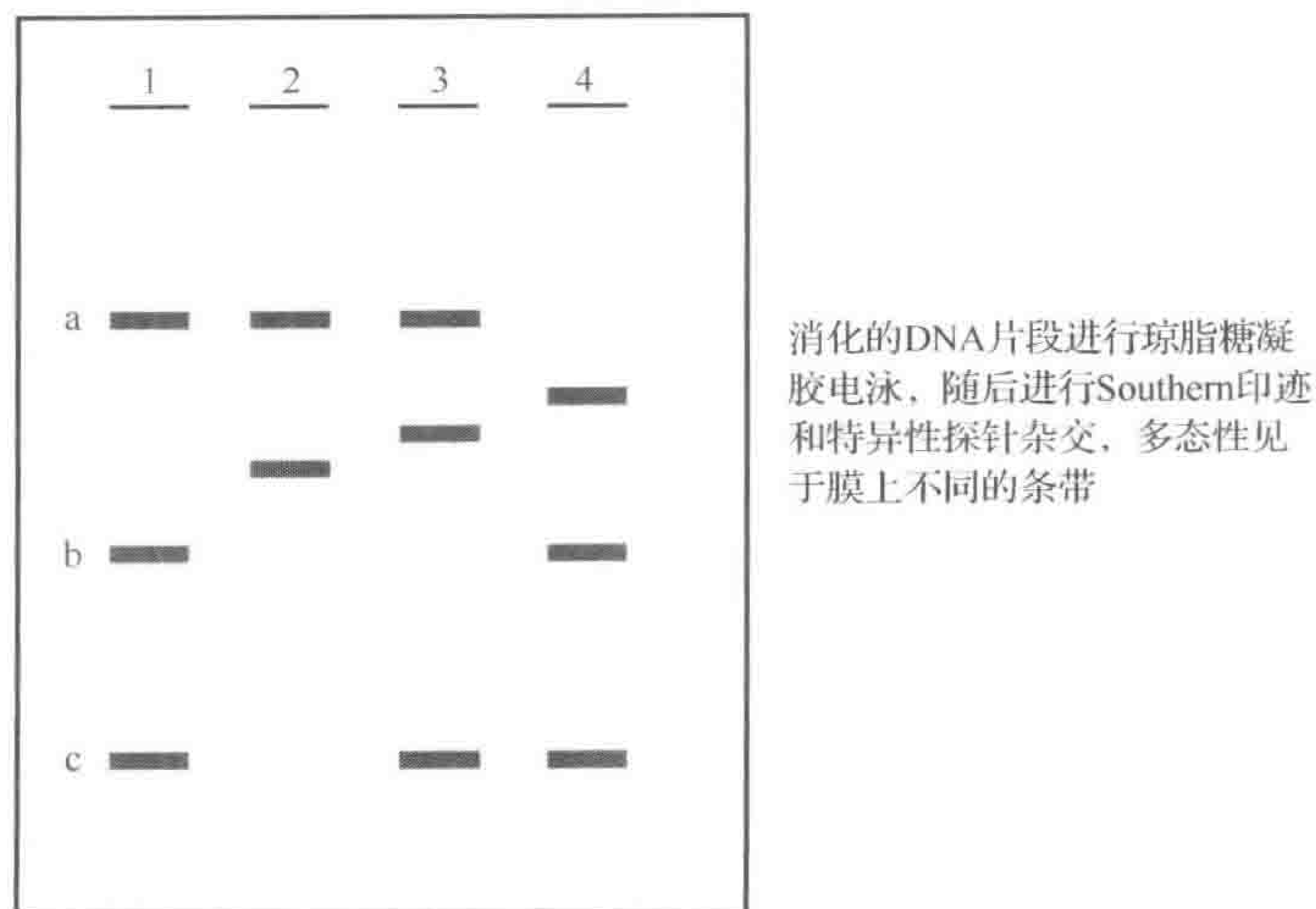


图 9.17 限制性片段长度多态性(RFLP)(续图)

该图显示染色体的区域和特异性探针进行杂交，在 Southern 印迹图中出现多态性的一些可能的原因：1. 最初的序列，探针检测标注的 a、b 和 c 3 种 DNA 片段；2. 通过突变引起的限制性位点的丢失导致片段 b 和 c 被单个较大的片段所取代；3. 在片段 b 内插入限制性位点改变它的大小，并没有改变片段的数目；4. 在片段 b 内删除限制性位点将减少它的大小

就此目的而言，插入序列是非常有用的，因为在一个菌株内经常存在多拷贝分子(在 Southern 杂交中呈现很多条带)，也因为其在染色体上的插入位点往往是高度可变的(引起广泛的多态性)。被广泛应用于流行病学调查目标的一个序列是结核分枝杆菌中 IS*6110* 插入序列。如图 9.18 中的箭头所示，使用这一探针，只有来自同一源头的菌株才会产生相似的模式。

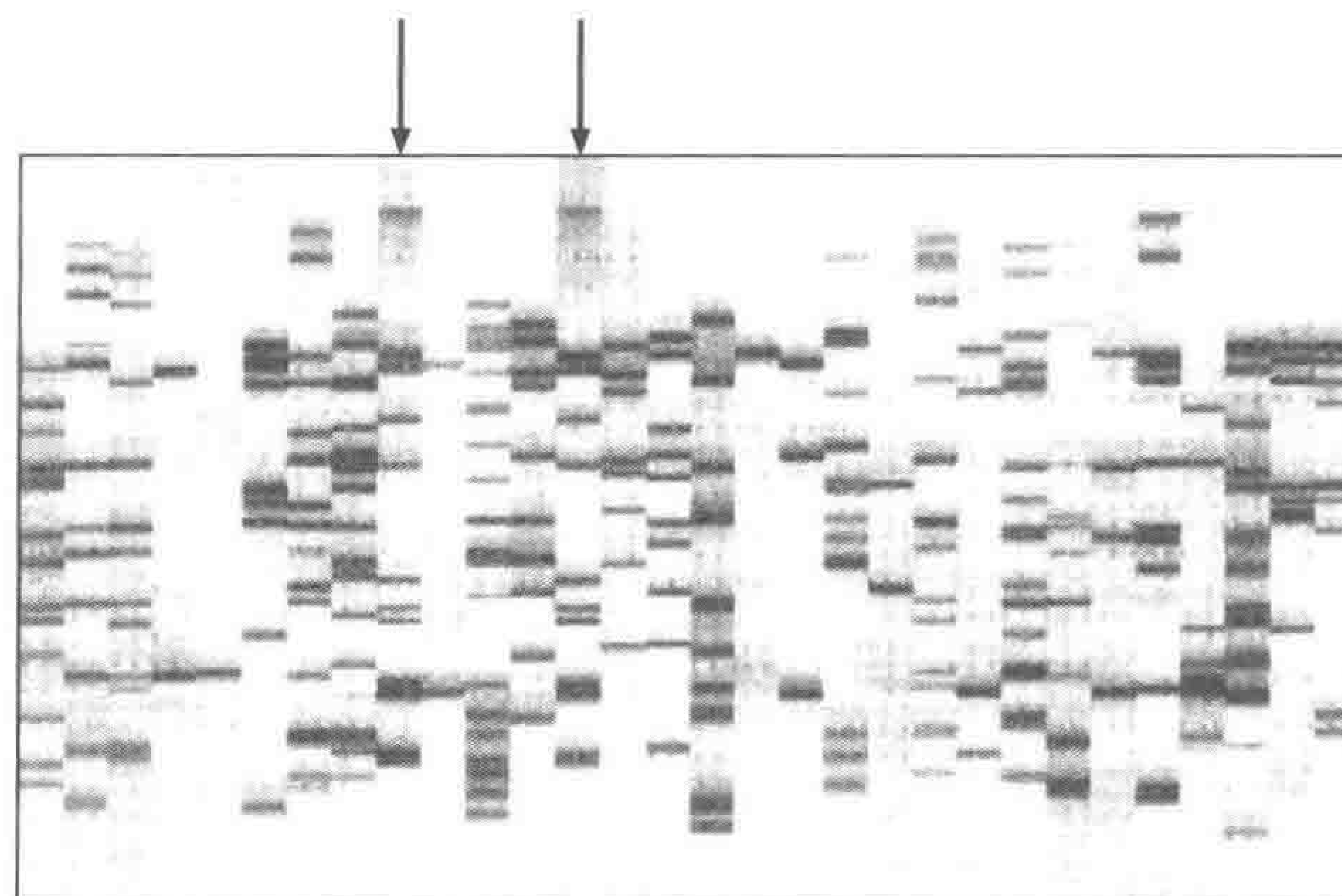


图 9.18 结核分枝杆菌的 IS*6110* 指纹法

从普通来源感染的分离病例显示相似的图形(箭头所示)

可变数目串联重复序列

另一种方法是使用不同类型的重复序列。细菌基因组通常包含串联重复序列，即方

向相同且没有任何干预序列的多拷贝重复的短序列(如 50～100bp)。这可能是由于滑链错配(见第 7 章),不同细菌的这些序列的每一个序列所在的特定位点的拷贝数各不相同。如果利用 PCR 技术扩增含有一个串联重复序列的 DNA 区域,可根据 PCR 产物的大小检测重复序列的拷贝数(图 9.19),这就是被称为 VNTR(可变数目串联重复序列)的分型方法。在任何一个位点的变异程度将是非常有限的,但基因组中通常有几个这样的基因位点,每个数据的综合便可产生很高程度的差异。

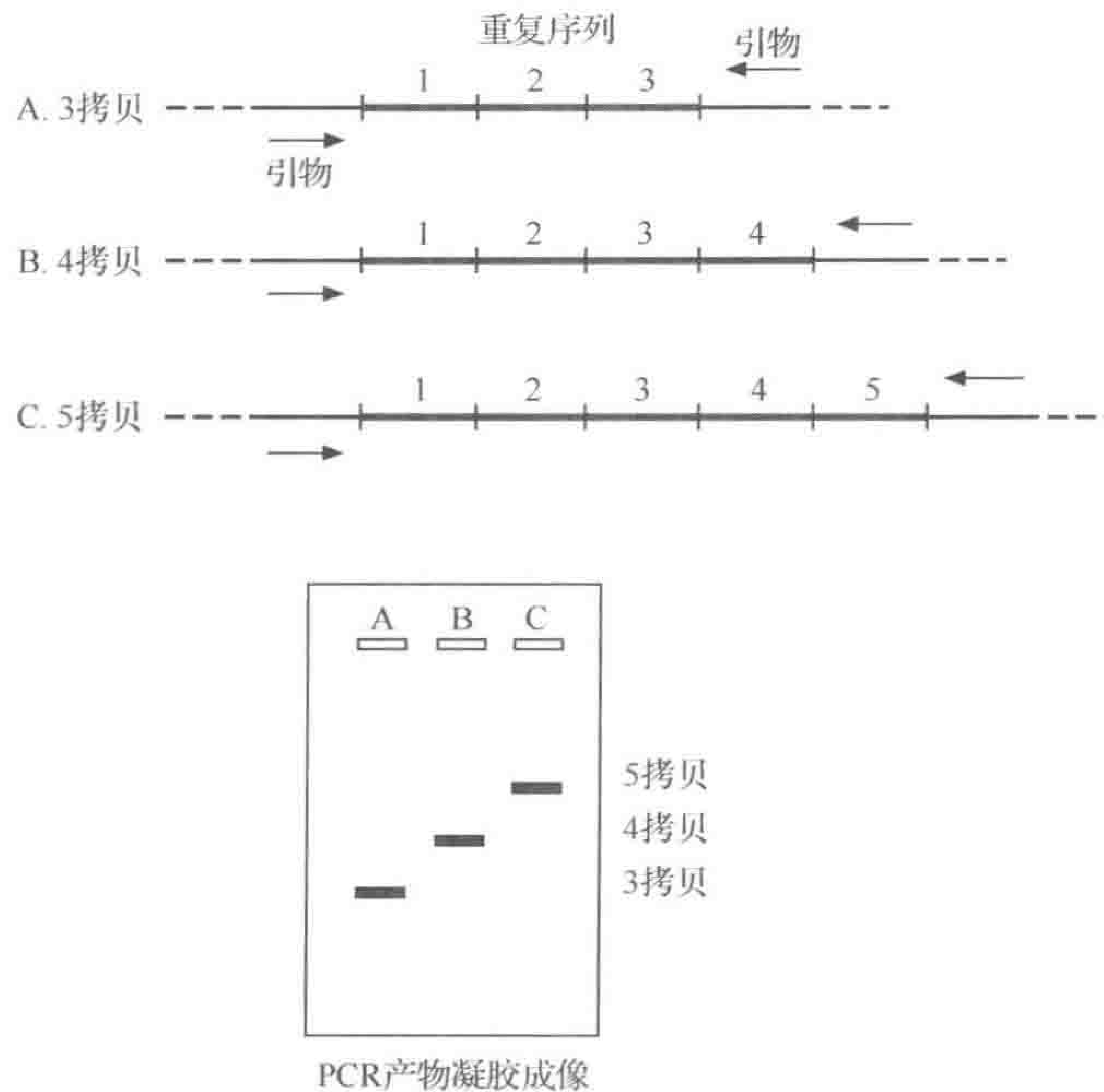


图 9.19 可变数目串联重复(VNTR)
包含串联重复的 PCR 扩增区域通过确定重复序列的拷贝数数目,可用于细菌的分型

RFLP 分型和 VNTR 方法也有一些缺点,它们只能检测基因组的相对小的区域。然而,通过与切割基因组 DNA 频率较低的常用酶的使用相结合,如 *Not* Ⅰ或 *Pac* Ⅰ(两种酶具有 8 个碱基识别位点,取代了其他限制酶识别的 6 或 4 个碱基对位点),RFLP 可提供全基因组的信息。当这些酶被用于消化基因组 DNA 时,可产生少量的大片段,如果能够解析,将提供相当有价值的跨基因组的分型信息。用于分离 DNA 大片段的技术经常涉及脉冲电场凝胶电泳(PFGE),并广泛用于细菌的分型,在第 10 章中将进行更详细的论述。

扩增片段长度多态性

扩增片段长度多态性(AFLP)是利用限制酶识别位点的变异区别菌株的另一种技术。因为它的重现性好且不需要任何事前的序列信息或另外的探针,所以它广泛用于鉴定密切相关细菌株的基因突变。最初的步骤和 RFLP 相似,即用限制酶消化来自感兴趣菌株的基因组 DNA。这一过程会产生数以千计的不同大小的 DNA 片段,在琼脂糖凝胶中,由于不容易解读这些条带,因此不具有信息性。AFLP 的基础是利用 PCR 扩增这些

片段并选择性地获得 100～300bp 的片段，因此可很容易地对不同菌株进行比较(图 9.20)。

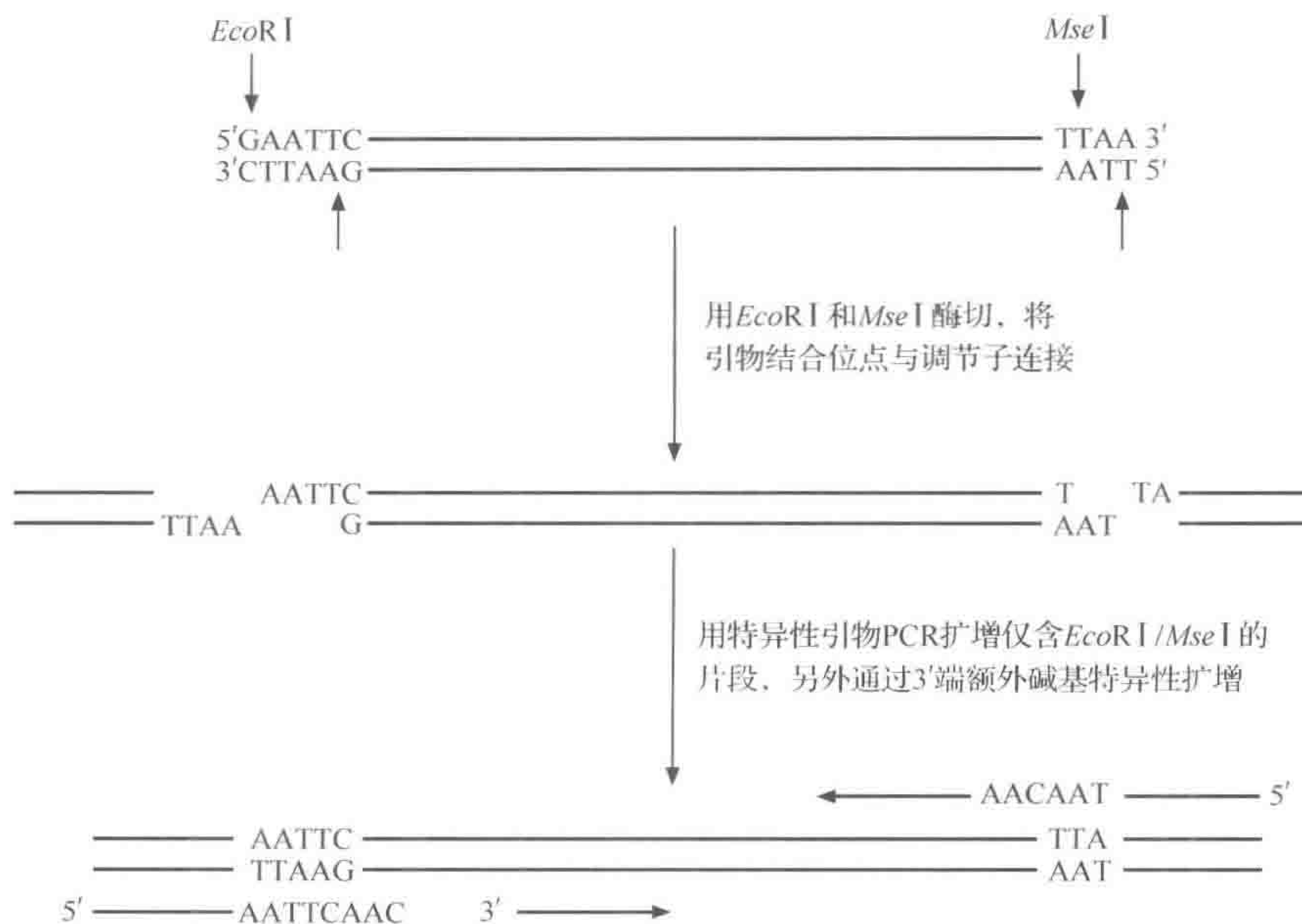


图 9.20　AFLP 的选择性扩增的原理

起始步骤是用两种限制酶对感兴趣菌株的基因组 DNA 进行酶切，一种限制酶通常具有一个 6bp 识别序列(如 *Eco*RⅠ)，另一种则具有 4bp 识别序列(如 *Mse*Ⅰ)。两种类型的转换接头，每一种都和限制酶切的突出端互补，同时也包括特异性 PCR 引物的靶序列，然后连接到酶切片段上。随后的 PCR 扩增步骤需要存在两个靶位点，故只有那些同时具备 *Mse*Ⅰ和 *Eco*RⅠ酶切末端的片段得到扩增。因此，用 *Eco*RⅠ/*Eco*RⅠ和 *Mse*Ⅰ/*Mse*Ⅰ酶切末端不能扩增，这些片段在随后的操作步骤中被排除，同时，在最后的可见操作中，不能检测到在开始步骤中产生的许多 DNA 片段。

在接下来的步骤中，加入与经转换接头引入的靶序列互补的两个 PCR 引物后，进行 PCR 反应。两种引物的 3'端区域与限制性位点靶序列互补，但延伸超过 3 个碱基。这一区域非常重要，因为它可以选择性地延伸，只有那些 DNA 片段恰好与该序列互补时才能得到扩增，所以当 PCR 反应进行时，仅部分合适的片段在可视化的操作步骤中被检出。常用荧光标记扩增片段，利用变性聚丙烯酰胺凝胶或毛细管电泳检测，因此，在对每一菌株的分析中，经常包括 100～300bp 片段，可以很容易进行比较。

多位点序列分型

在限制酶识别位点位置差异的基础上，RFLP 和 AFLP 进行基因组比较。虽然二者都广泛应用且在菌株比较中非常有用，但两种技术均存在一个潜在的限制性，即它们并不揭示变异序列发生的位置，以及突变的发生是由于突变积累还是由于重组。如果菌株存在的差异是由突变积累所引起的，则两种菌株是相关的，并能追溯到共同的祖先，因此这

在追踪菌株的进化途径时尤为重要。然而,如果这种差异是由重组所引起的,即使这两种菌株在此位点及基因组上的这部分序列均具有相似性,但也可能是不相关的。

多位点序列分型(MLST)是用于鉴别密切相关菌株的另一种技术,该技术通过大量分析基因组不同部位的、较长片段的DNA序列的差别,能够提供明确的流行病学信息。MLST的原理非常简单,它包括所选基因组区域的PCR扩增及随后的PCR产物的序列测定,然后对序列进行比较分析。那些细菌细胞中编码执行基本和必需功能蛋白的管家基因,因为它们在菌株间高度保守并且随着时间延长发生突变积累的速度很慢,通常被选作该技术的靶基因,如编码甘油激酶和乙酰辅酶A转乙酰酶的基因。

对于每一菌株,可以测定多个基因(通常是7个片段,可依据所需的鉴别程度增加或减少)内部片段(每个片段的长度通常是450～500bp)的DNA序列。每一特定基因的不同序列突变体被认为是独特的等位基因,而且每个独立的各个不同位点的等位基因被定义为一个特定菌株等位基因的分布或序列分型(ST)。虽然MLST分析比RFLP和AFLP更耗费时间,但由于它从基因组上多个位点获取信息,比这两种技术更适合大范围的全面流行病学调查。另外,由于管家基因的突变是一个相对慢的过程,细菌株的等位基因分布随着时间的推移更趋于相对地稳定。

考虑到以上几种情况,这些分型方法的应用可延伸至细菌株进化的研究。为了达到这一目的,需要考虑分型方法是否能成功地鉴定相关的菌株。例如,如果鉴定一种特定的RFLP或AFLP型,是否也能预测菌株的其他特性?如果细菌种是纯克隆的,例如,菌株(克隆)的所有成员来源于单一个体,没有水平基因转移,我们认为哪一菌株的成员彼此比其他菌株更相近。因此,希望该分型方法可预测生物体的其他特性。有时确实发生过这种现象,例如,*E. coli* O157血清型与毒力程度相关。事实上,许多微生物诊断方法是建立在这些相关的基础上的。例如,应用凝固酶检测区分致病性金黄色葡萄球菌(*Staphylococcus epidermidis*)和低致病性的葡萄球菌,如表皮葡萄球菌(*S. epidermidis*)。然而,必须说明的是,毒力和其他显著的特征通常太复杂,不适合这样简单的分析,尤其是会受到宿主反应及生物体典型特征的影响,这些目前尚处于争论中。

另外,如果发生水平基因转移,这一菌株将不是克隆的。在这种情况下,这种相关性将不再存在。事实上,不同分型方法的结果可能彼此并不吻合。因此,当用独立的方法验证由一种分型方法得出的明显的进化关系时,其结果明显不同。

10 基因组的基因定位及其他

本章主要回顾了研究基因定位、结构和表达的方法。以前这主要是通过大量使用在体内条件下的基因定位完成，但由于 DNA 测序和其他技术的迅速发展及其应用，大大推动了其进程。因此，本章也充分考虑了这些使全基因组范围的基因研究成为常规的研究方法，也将描述一条基因组序列如何呈现有意义的信息以及研究单条基因和基因组范围的表达的方法。本书对现代分子生物学和遗传学中最令人振奋的内容只做简单的介绍。

10.1 基因定位

经典遗传分析的主要目标之一是确定基因在染色体上的位置。这项工作看似晦涩难解，但系统了解染色体相关知识的确在理解基因功能中起到重要作用，并已在第 9 章中对其进展进行了全面深入地描述。

细菌基因定位的经典方法依赖于应用接合、转化、转导，通过基因转移产生的重组产物，尽管其目前已被基于体外基因技术的方法大量取代，但对这些方法的基本理解对于细菌遗传学认识极具价值。

10.1.1 接合分析

在 6.2.2 部分，已经了解到 F 质粒整合到大肠杆菌(*E. coli*)染色体可产生能转移 1 个拷贝的染色体到适宜受体的 Hfr 株。转移整个染色体需要 100min。鉴于此，*E. coli* 遗传图(图 10.1)可从 0～100min 进行标记，按顺时针方向转移，而且每个基因被设计一个位置，这一位置与其从苏氨酸基因座的任一起点转移的时间对应(*thr*，0min)。由于 F 质粒可整合到不同的位点和(或)以不同的方向，转移特定基因的实际时间、转移的方向均依赖于所使用的 Hfr 株。

然而，整个染色体的转移是很罕见的。配对的碱基对将变为独立的、随机分布多次。转移一个基因需要的时间越长，在转移之前，配对碱基对将有更多的机会分离。这可能由于转移梯度与转移起始点的基因位置有关。

有一个便捷的方法用来确定 *E. coli* 染色体上基因的位置(图 10.1)。如果一个原质营养的 Hfr 株与一种增殖的营养缺陷型(如 *thr*、*leu*、*trp*、*his*、*arg*)受体菌交配，通过将混合物的一部分接种到添加了 4/5 种氨基酸的最低限度培养基，如缺乏 thr，可以确定已接受了每种标记的受体菌数量。这是阻断原质营养供体生长所必需的，如在培养基中存在链霉素，而使用链霉素抗性的受体菌。这里链霉素被用作反选择试剂。在这种培养基中，供体菌不能生长(存在链霉素)，亲本的受体菌也不能生长(不存在 thr)，只有重组的接受

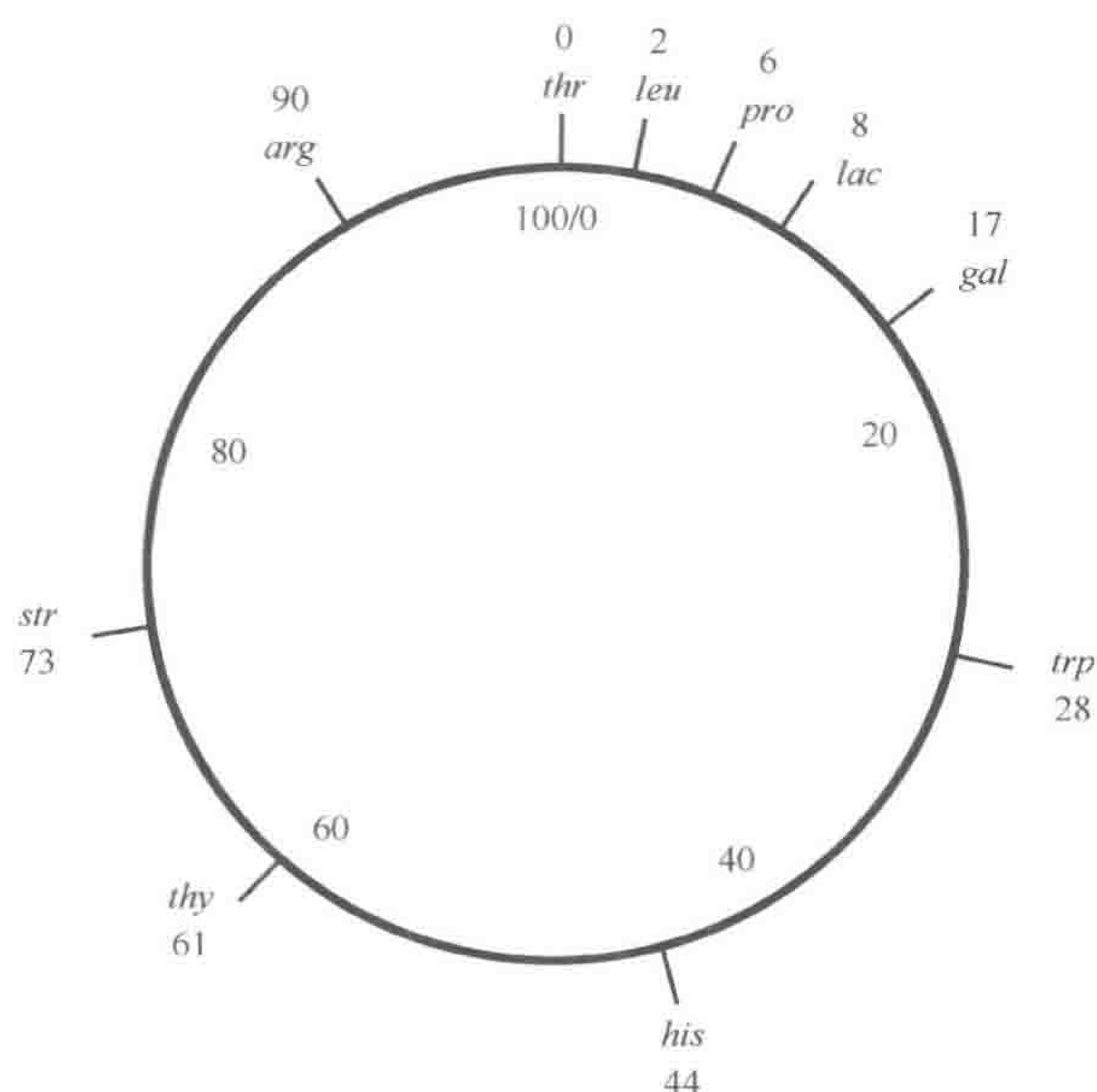


图 10.1　大肠杆菌 *E. coli* 遗传图

显示的基因是合成苏氨酸、亮氨酸、脯氨酸、组氨酸、色氨酸、精氨酸等的位点，利用乳糖和半乳糖的位点，以及链霉素抗性的位点。在许多情况下，标注出这些位点表明是基因簇而不仅仅是单个基因。精氨酸调节子也包含了其他位置的基因。图谱时间的数字代表了在苏氨酸位点开始的接合转移所需的时间(min)。*thr*，苏氨酸；*leu*，亮氨酸；*pro*，脯氨酸；*lac*，乳糖；*gal*，半乳糖；*trp*，色氨酸；*his*，组氨酸；*thy*，胸腺嘧啶；*str*，链球菌素链霉素；*arg*，精氨酸

thr 基因的受体菌才能生长。

这种结果可从图 10.2 中得到说明。在此例中，使用了 Hfr 供体，基因从非常接近于 *thr* 基因座的位置以顺时针方向从 Hfr 供体转移。重组体数量的对数值和相关基因的图谱位置之间具有线性关系。如果不知道 *trp* 基因的位置，通过确定 Trp^+ 重组菌的数量可以如图 10.2 所示，将基因在图谱中定位。

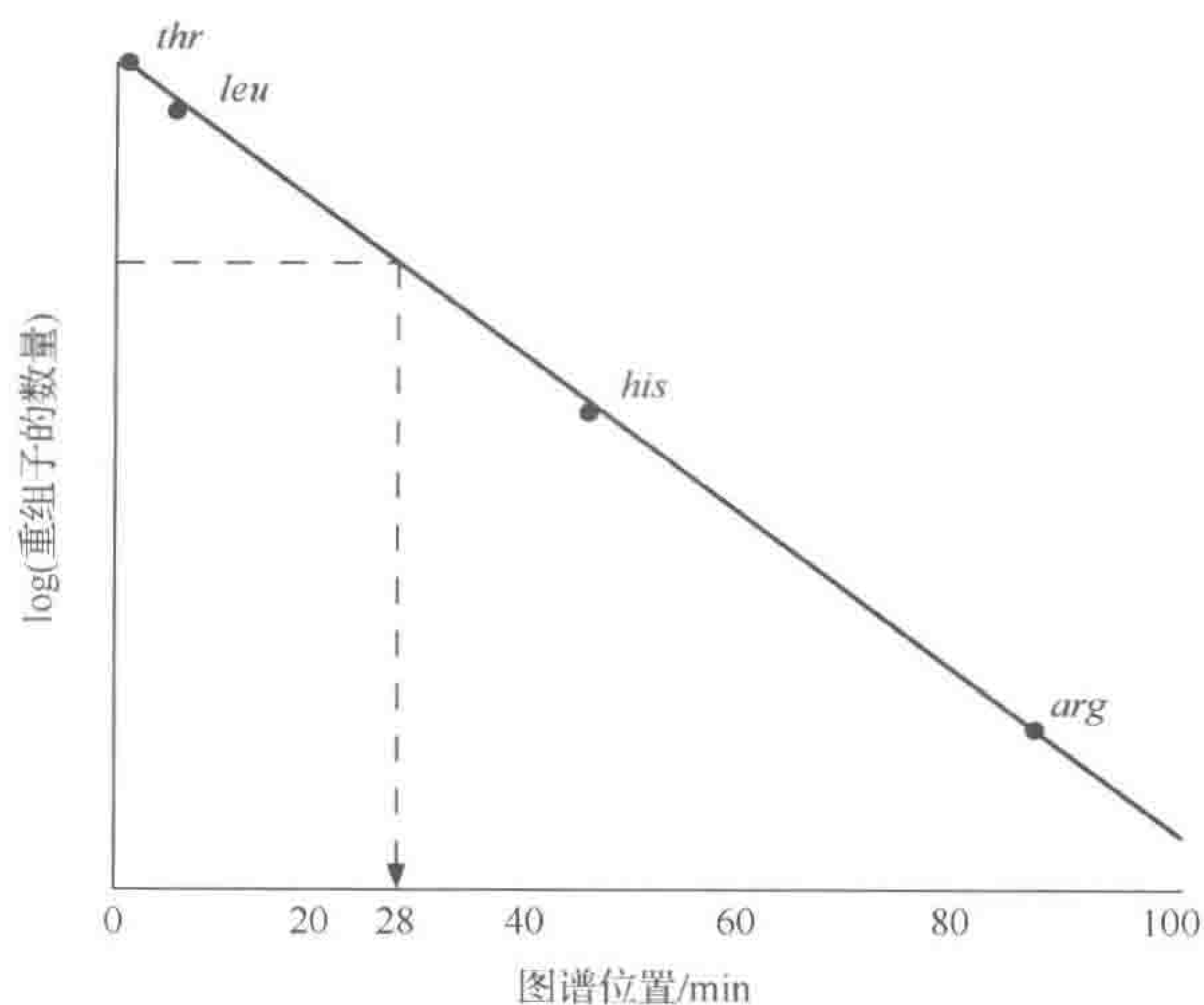


图 10.2　使用接合转移梯度的遗传图

thr，苏氨酸；*leu*，亮氨酸；*his*，组氨酸；*arg*，精氨酸

10.1.2 基因文库

基因克隆技术为经典基因作图技术提供了一个强有力的选择。由于基因组文库中的每个克隆均携带有限数量DNA，可以用于鉴定染色体上相邻的两个基因，如同时检测两个突变体的互补能力。

更为普遍的是，可由杂交方法来确定文库中基因的相对位置，使用特异探针（图10.3中探针1），如果从挑取含有第一个基因的克隆开始，用这个克隆的插入片段作为探针鉴定含有基因组重叠区的第二个克隆，这第二个克隆可以作为探针进一步鉴定含有重叠区的克隆，直到获得感兴趣的第 n 个克隆，这个过程称为染色体步移，也可用于已知两个基因位置的作图或逆向克隆一条与已鉴定标记相近的未知基因。

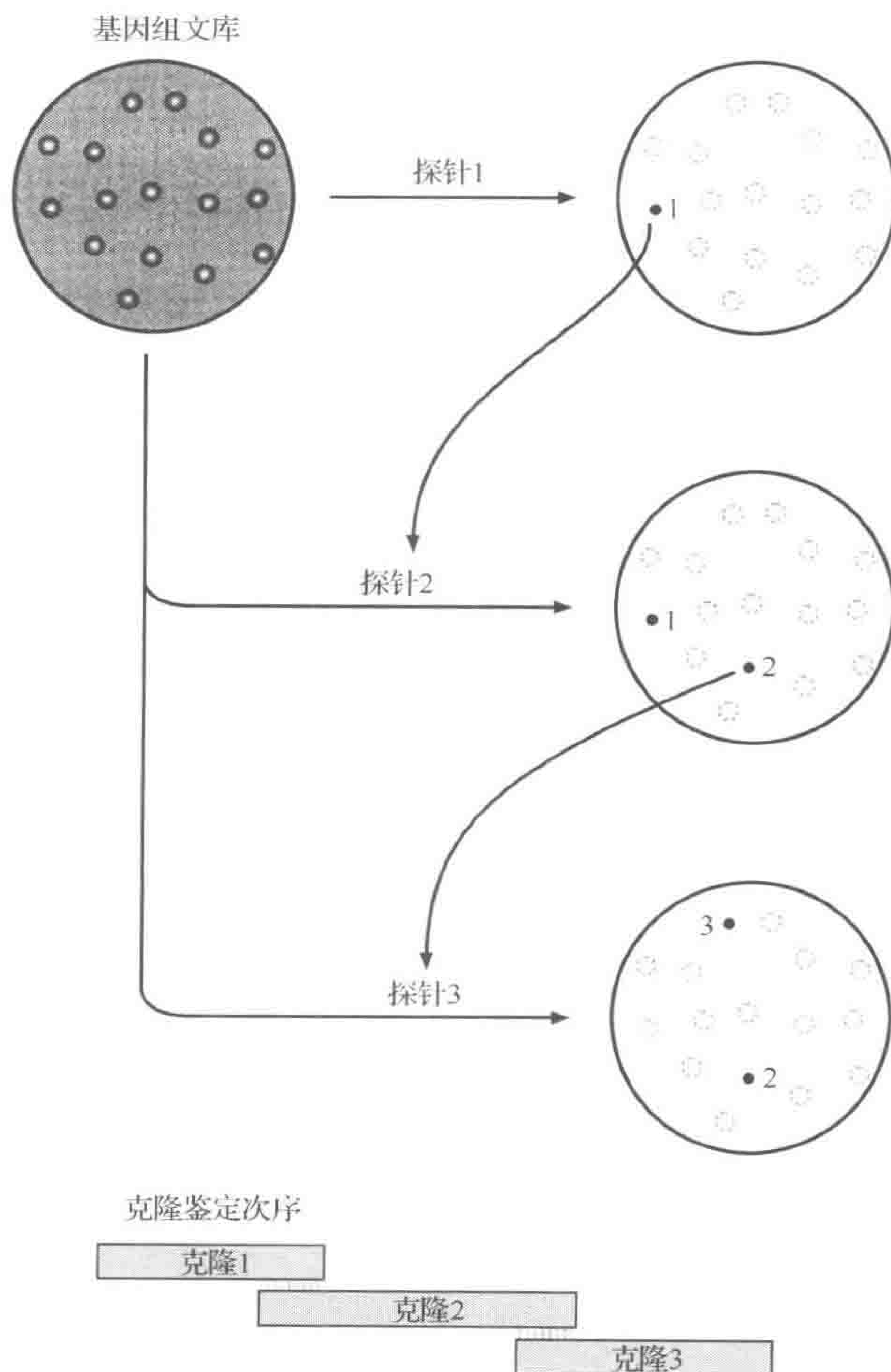


图10.3 染色体步移和有序文库的构建

克隆1，通过使用基因组文库中的探针1而确定，并被使用产生探针2重新筛选文库，鉴定出重叠克隆（克隆2），并用于鉴定克隆3，以此类推

此项技术可拓展应用于获得完整重叠克隆，这些克隆以其在染色体上的序列有序地进行排列。这又称为有序文库，构建一个有序文库需要大量的工作，可是一旦建立，将是一个非常有价值的资源。一些基因组测序计划已有效应用了有序基因文库。

10.1.3　限制作图和脉冲场电凝胶电泳

通过限制酶单酶切或多酶切 DNA 片段，分析酶切片段的大小，可以确定各限制位点的位置。限制作图常作为一个关键步骤来鉴定 DNA 克隆片段，现在通过 DNA 测序更易做到。然而，限制位点或片段大小分析仍然在应用，如不同菌株染色体结构的比较，但如果使用很多限制酶消化细菌总 DNA，会得到一些在传统凝胶电泳中不易处理的片段，通常出现弥散条带(smear)。在第 9 章中，通过特异性探针进行 Southern 印迹用于菌株 RFLP 分型，或者使用酶切片段少的酶，如 *Not*I 或 *Pac*I(均有 8bp 识别位点，而常用限制酶的识别位点为 6bp 或 4bp)，这类限制酶在大片段上识别位点数量相对较少，因此可能产生数量较少的大片段。

然而，大于一定大小的线性 DNA(通常约 20kb，但随条件改变而改变)在传统琼脂糖凝胶上不能分离，片段可在胶上迁移，但以不依赖片段大小的方式。大 DNA 分子迁移时会两端相遇，一旦排在一起，其长度就不重要了。这与那些随机构象足够小的小片段可通过胶孔相反。

这些大片段可在不断改变电场方向的凝胶中分离(图 10.4)。所有的片段实际以同样的速度在胶中迁移，但由于电场方向的改变，它们不得不在新方向下转变方向。分子越大，转向时间越长。结果是在胶的起点附近分离到较大的分子。

这项技术称为脉冲电凝胶电泳(pulsed-field gel electrophoresis, PFGE)，包括以下几种类型。最简单但具有局限性的方法是反转电场凝胶电泳(field-inversion gel electrophoresis, FIGE)，常使用一个带有短脉冲反向极性电极的普通电泳槽。更有效的技术依赖于电极的混合排列，应用最广泛的系统之一是使用六角形排列的电极的高钳位均匀电场凝胶电泳(contour clamped homogeneous electic field gel electrophoresis, CHEF)，控制其电场强度和方向，能够产生均一的电场，确保 DNA 分子按直线迁移。

10.2　DNA 序列分析

分子生物学的主要目标之一是了解 DNA 的结构，以及信息是如何翻译成细胞内复杂的工作方式的。显而易见，如果能确定 DNA 的碱基序列，这项任务则可能会完成。1977 年，Frad Sanger 研究出一种酶链末端测定方法使之成为可行，首次能一次测定 DNA 链序列 100bp。目前，DNA 测序技术已经达到一个顶峰，一次运行几个小时就可以完成可覆盖细菌基因组草图序列的 40～60 亿 bp 的测定。

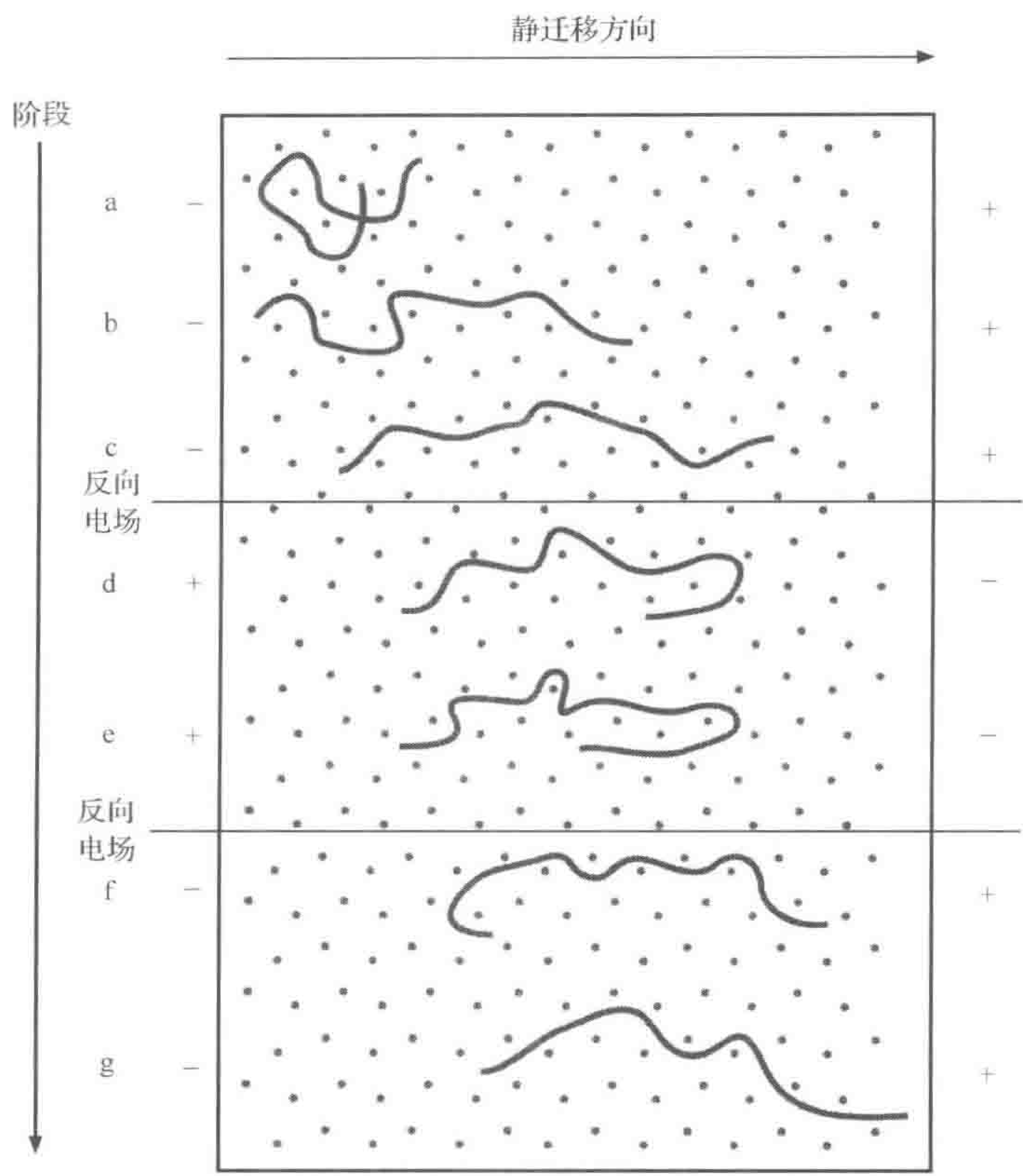


图 10.4 脉冲场电场凝胶电泳的原理

(a) DNA 分子太大不能在胶中以最终状态迁移。(b、c)在电场影响下，慢慢适应而能移动，并随电场方向而自动调节定位，因此可以迁移。(d)当电场的方向改变，(e)它开始转换方向。(f)在未完全转换完方向时，电场又改变方向。(g)分子不得不在开始移动前又从左向右改变方向。一旦分子确定方向，它的移动不依赖于大小，决定迁移率的主要因素(脉冲环境)将决定转向的快慢，这是片段大小的主要功能

10.2.1 Sanger 测序法

Sanger 测序法是研究一种体外 DNA 聚合酶和 2′,3′-双脱氧核苷三磷酸(dideoxynucleotides，ddNTP)参与的 DNA 合成技术。由于 ddNTP 缺乏延伸所需要的 3′-OH 基团，故不能与延伸的残基相配对，使延长的 DNA 链在某处终止，而不能延伸 DNA 链。

新链的合成始于特定位点，通过使用一条引物，其互补 3′端附近的位置可以被测序(图 10.5)。如果每个反应含有 4 种脱氧核苷三磷酸(dNTP)中的一种(如 ddATP)，这将使得每个应加有 A 碱基的位点都有加入 ddA 的机会，这将导致 DNA 链的混合物的每一种都会在 A 碱基终止。由于其长度不同，可以用丙烯酰胺凝胶分离，并可以表明每个 A 端的位置。依此类推，可以通过掺入 4 种 ddNTP 中的一种，从而出现图 10.5 的胶图谱条带，从中可以读出序列。在一块胶中可以读出几百个碱基。



随着dATP和ddATP的出现，将出现由ddATP末端终止的混合片段



由ddGTP、ddCTP和ddTTP介导的相似反应，来自4个反应的片段在聚丙烯酰胺凝胶上分离，并被自动放射仪检测

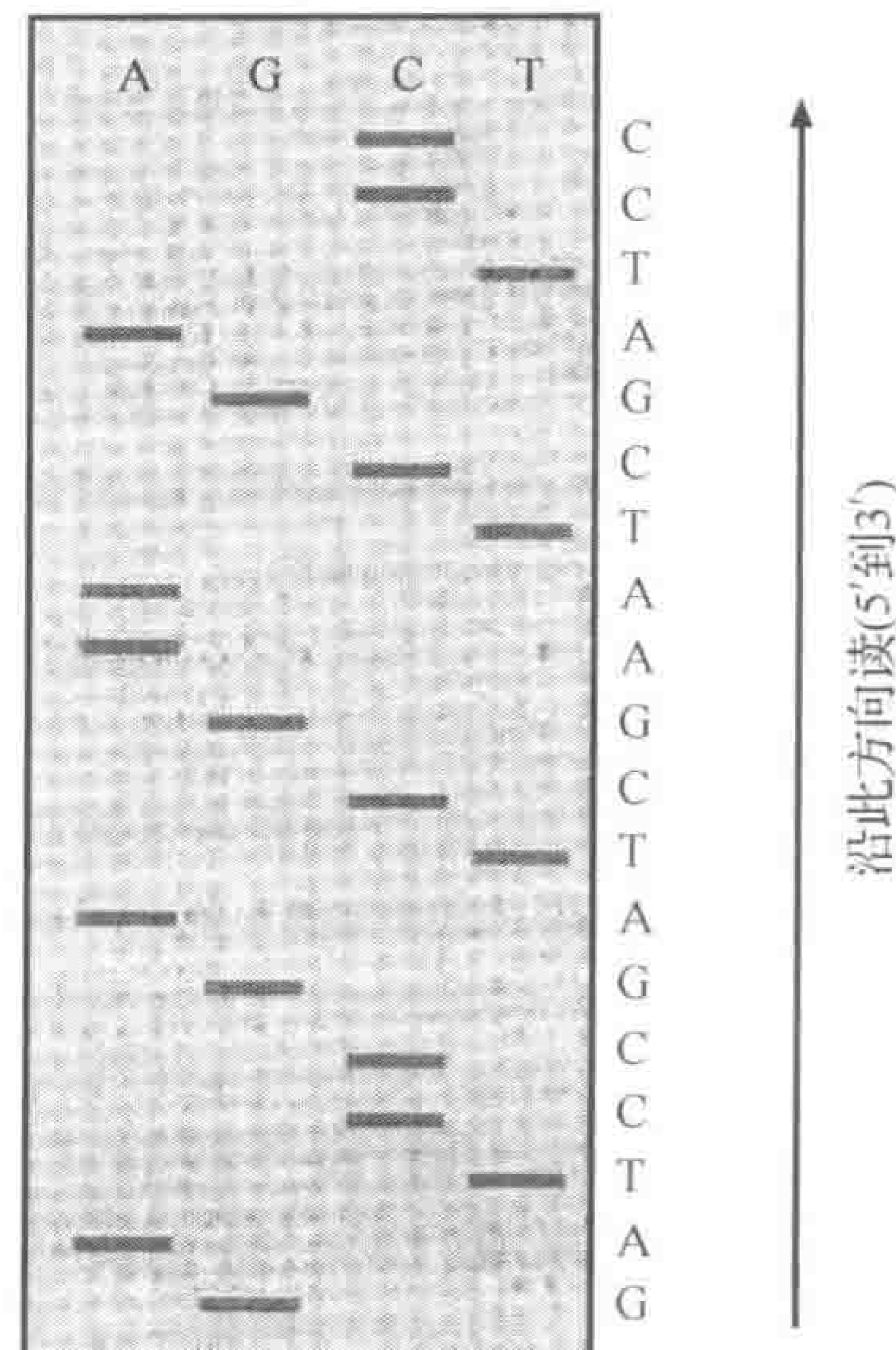


图 10.5　DNA 测序策略

为获得较长的序列，通用的策略(如鸟枪法测序)是随机切割 DNA、重叠片段、克隆，然后对每一个片段进行测序(图 10.6)。由于每个新片段被测序，通过计算机比较所有以前的序列，当发现一个重叠时，计算机加入两个序列为一个集合，当更多的序列被测定，集合越来越大，并与其他集合重叠，重叠的集合形成更大的集合，最终形成覆盖整个基因的单集合。

通过使用计算机产生翻译产物的序列，基因测序技术的发展也为蛋白质序列的推导提供了一条途径。在这之前，蛋白质序列的确定是一项昂贵而艰巨的任务。现在由于测定基因比测定蛋白质容易，此项技术极大地推动了蛋白质生物化学进程。虽然 Sanger 人工测序方法需用几天或数周的时间才能得到可靠的 200～300bp 序列，经过数年才测定出单个基因的序列，但这是分子生物学上一个重要的突破，因此，Sanger 获得了诺贝尔化学奖。

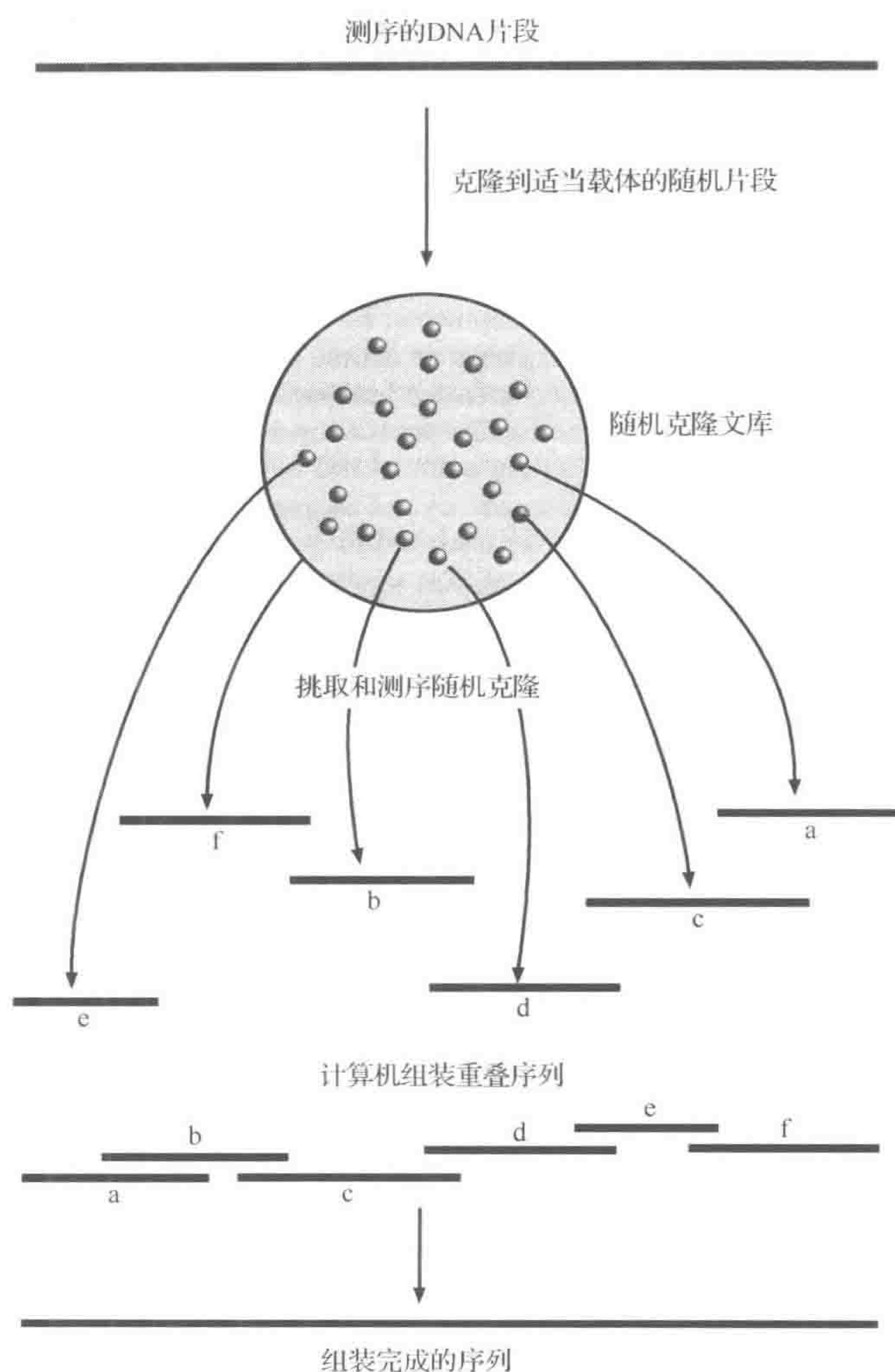


图 10.6　鸟枪法克隆和测序

10.2.2　染料终止法测序

Sanger 的方法尽管是一项伟大的发明，但也是非常费时的工作，一次只能测几百个碱基而且需用肉眼辨定结果。1986 年，为了使之更快，自动化程度更高，Leroy Hood 将 Sanger 的过程简化，去除了放射性标记，并用不同的荧光染料分别标记每种链末端 ddNTP。如上所述，每种末端核酸被标记为红、绿、黄或蓝 4 种不同的颜色，DNA 片段由于其掺入而表现不同的颜色。

因此这 4 种测序产物可混合于同一反应管中，可直接上样于聚丙烯酰胺胶的同一泳道中，这显著提高了处理量。另外，自动测序仪的设计结合了可读取胶中样本的激光仪和可分析结果的计算机，这些早期的设备迅速推进了测序量，可提高到每天测 4800bp，并促

进了更多宏大的测序项目的开展。一些现代的 DNA 测序仪仍然使用这种化学技术，但聚丙烯酰胺胶梳子已被毛细管矩阵所替代，每管相当于老式胶的泳道。这样的设备能平行运行 96 个毛细管，每天测定 200 万个碱基。这些 DNA 测序仪仍是大多数分子生物学实验室的中坚力量，随着新测序技术的飞速发展，它们已被新的测序技术所取代，主要有如下几种。

10.2.3 焦磷酸测序

尽管测序技术已经取得了巨大进展，科学家仍然在努力研究出更快速、简便、价廉的方法。1998 年，推出一项全新的测序技术，该技术无须进行电泳，核酸和引物也无须荧光标记，操作更为高效和快捷。就像上面描述的两种测序技术一样，此项技术被命名为焦磷酸测序(pyrosequencing)，通过合成互补链，加快了 DNA 的单链测序。最大的不同是焦磷酸测序基于在核酸掺入时检测焦磷酸的释放，而不是使用链终止。当合成一条 DNA 链时，每次只在延伸的单链分子中加入一种底物 dNTP，同时释放出一个分子的焦磷酸(PPi)，并促进合成反应的进行。焦磷酸测序依赖于每次一个正确的碱基掺入到新合成的 DNA 链中所释放焦磷酸的检测。从其自身来说，PPi 是不容易被检测到的，因此，检测过程有赖于将 PPi 的量转化为发光检测(酶的级联反应，图 10.7)。在这种方式下，在 ATP 硫酸化酶的作用下，生成的 PPi 可以生成 ATP，生成的 ATP 为萤光素酶提供能量，萤光素酶氧化萤光素，产生可见光。

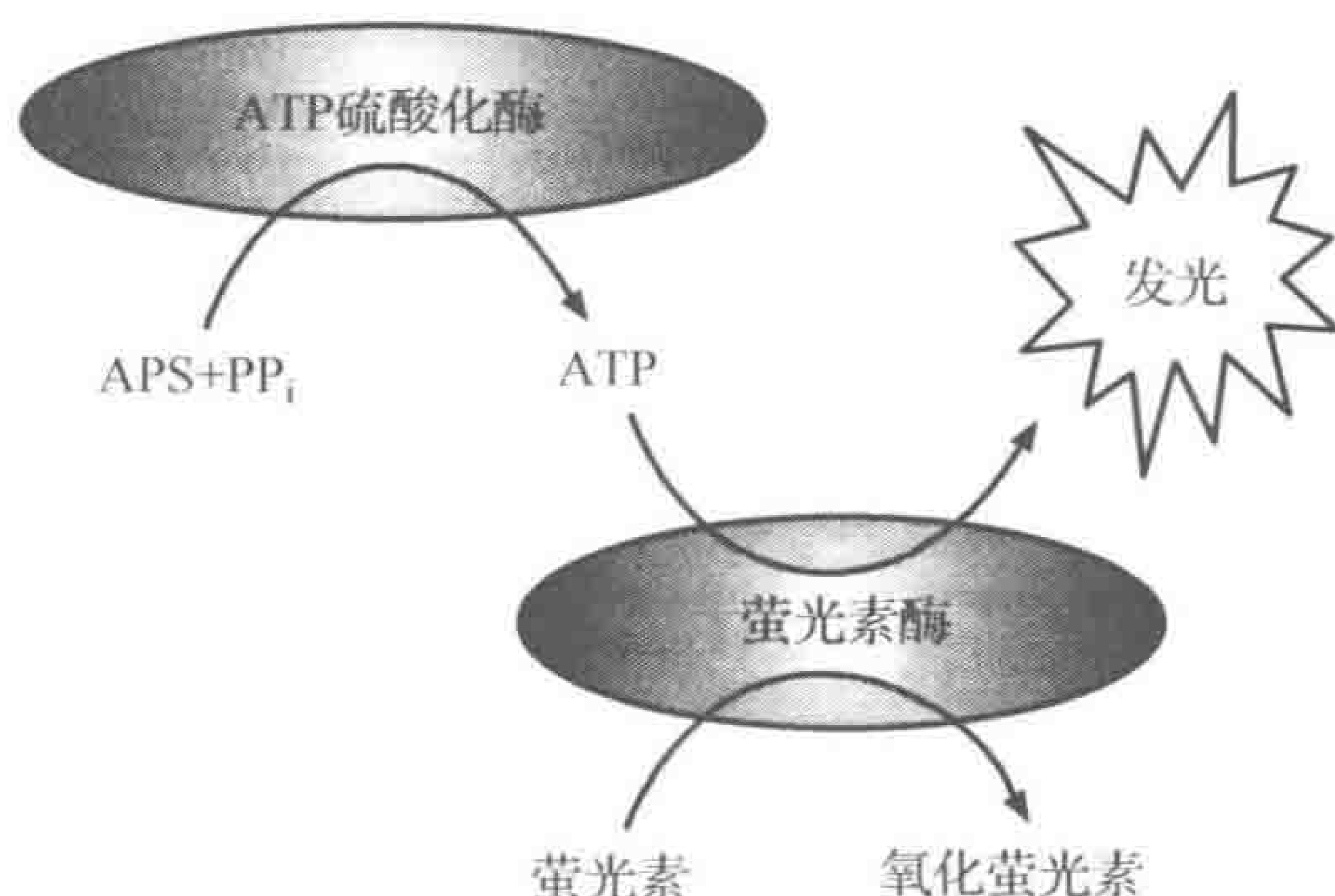


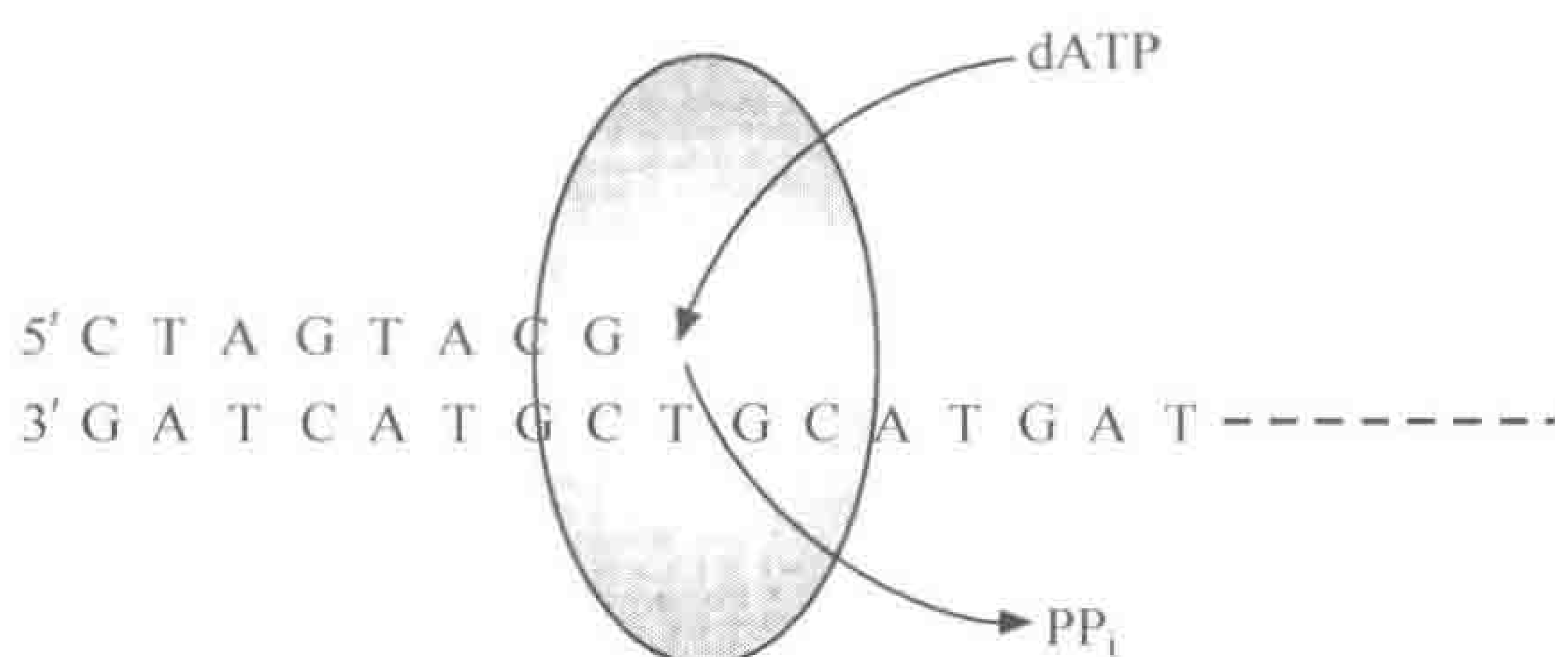
图 10.7 焦磷酸测序：焦磷酸(adenosine 5′-phosphosulphate，APS)的检测

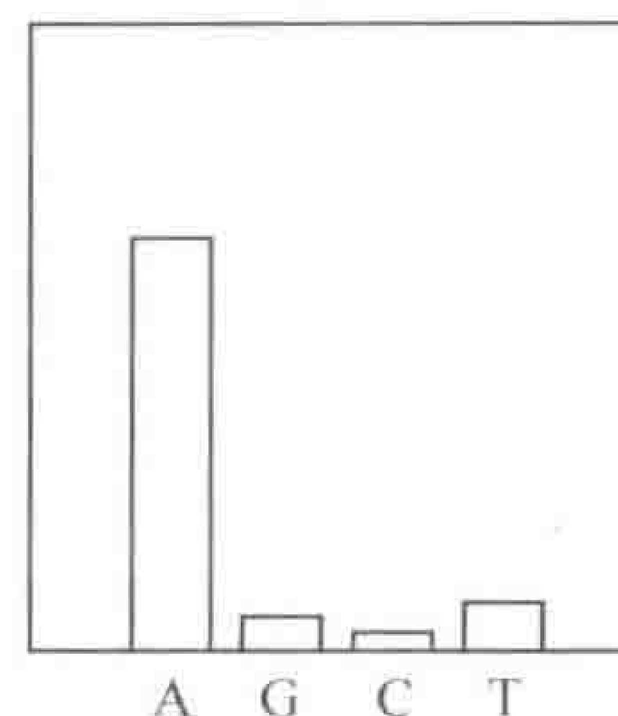
在实际测序反应中(图 10.8)，模板 DNA 被固定，加入引物、DNA 聚合酶，然后依次单独添加 A、C、G、T 核苷酸溶液。每次添加之间要进行严格的清洗。清洗缓冲液包含可以水解 ATP 和其他残留核苷酸的酶——三磷酸腺苷双磷酸酶，以便在下一次添加前消除可能的污染。加入一种核苷酸后发光，表明它一定与模板的下一个未配对碱基互补。

10.2.4 大规模平行测序

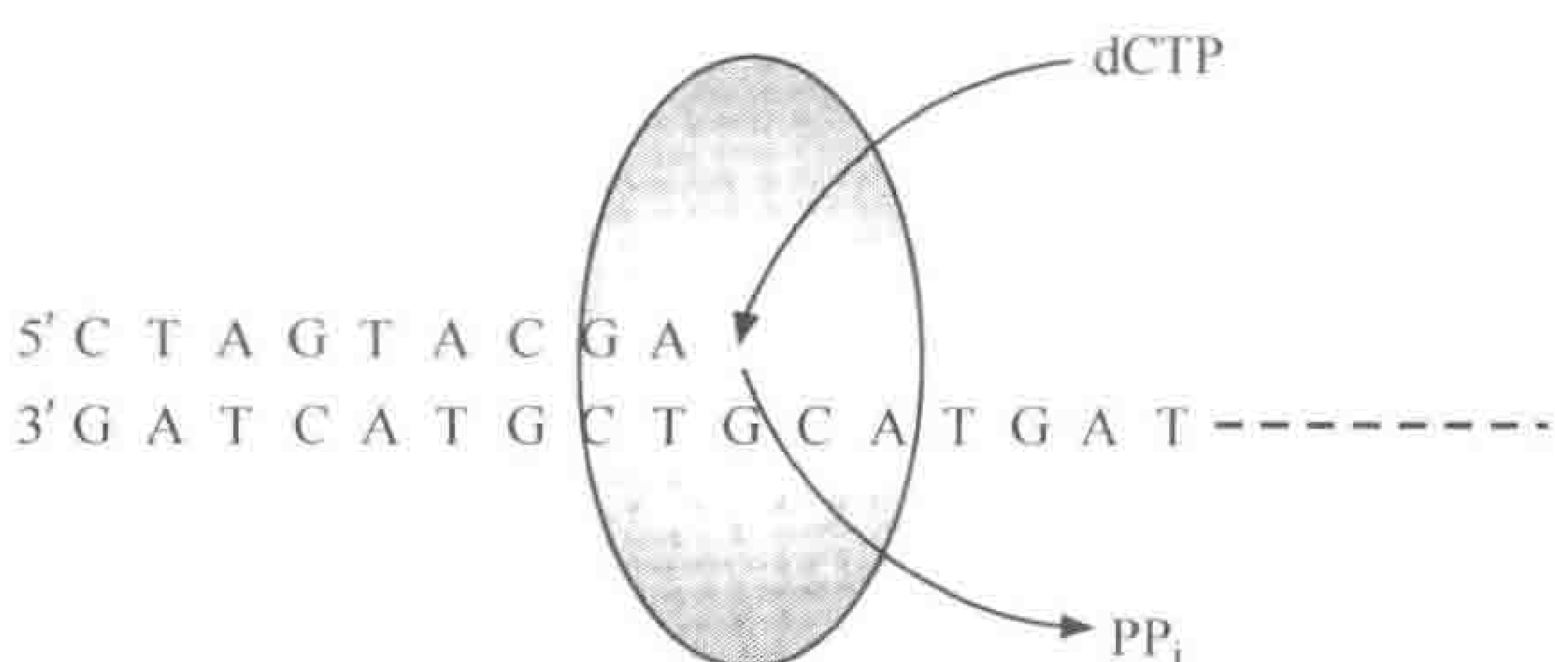
就其本身而言，焦磷酸测序不能被称为是一个重大的飞跃。然而其在自动化方面的进一步推进，双脱氧测序是不可能的。这一进步称为大规模平行测序，由于获得 DNA 序

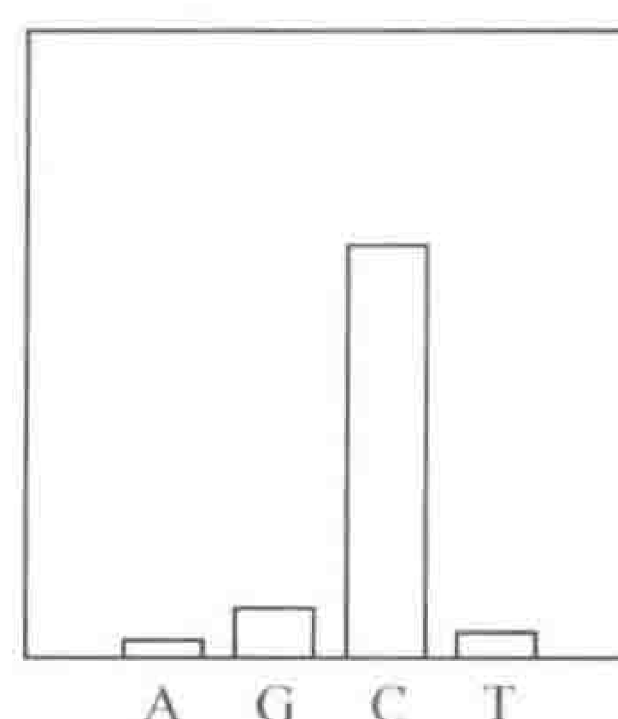
步骤1. 加入dATP, 与其他dNTP无反应



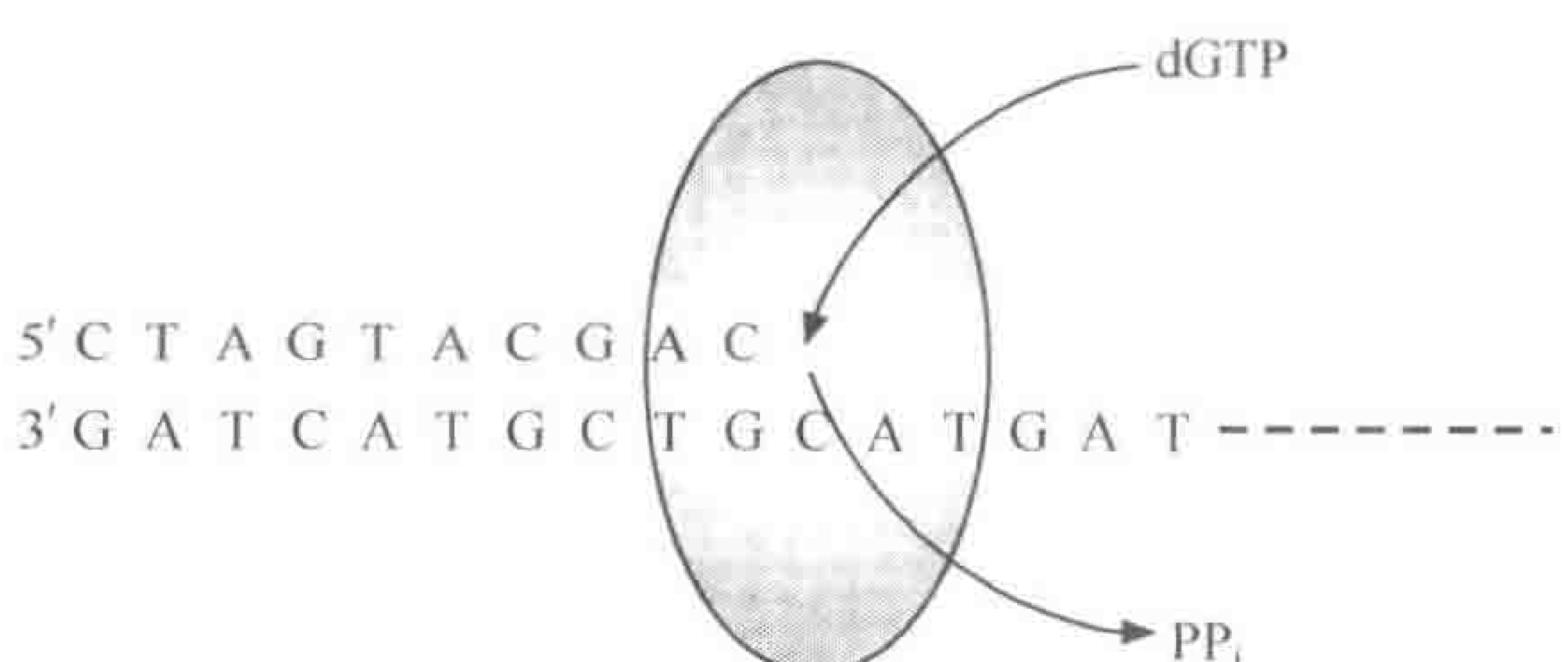


步骤2. 腺苷三磷酸双磷酸酶法，加入dCTP, 与其他dNTP无反应





步骤3. 腺苷三磷酸双磷酸酶法，加入dGTP, 与其他dNTP无反应



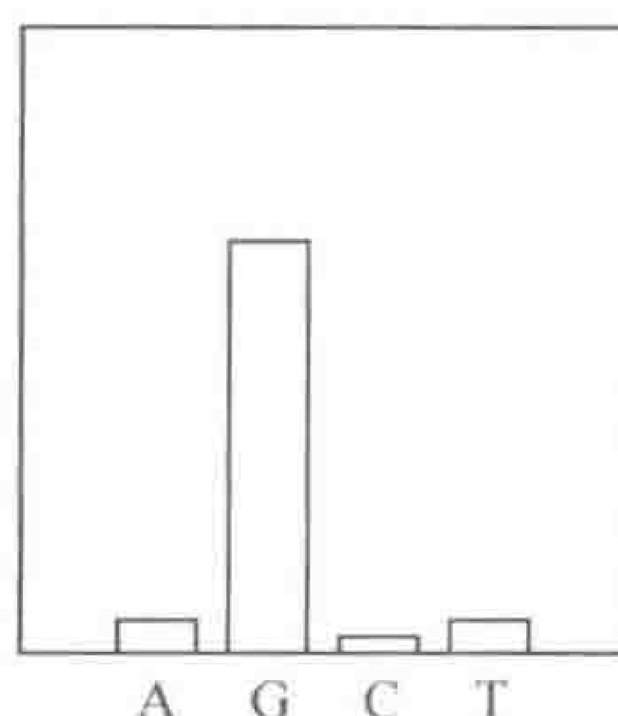


图 10.8 焦磷酸测序

列的速度要比现有技术测序系统快 100 倍，已为 DNA 测序提供了新的方法。通过提高能够同时完成测序反应数，该系统大大提高了产量。

DNA 测序从基因组 DNA 被打断为小片段开始，接头序列可将这些片段连接，通过接头序列，DNA 缠绕于直径约为 28μm 的微珠上，每个微珠上仅缠绕 1 套基因组 DNA，有数百万个缠绕 DNA 的微珠。在测序前，DNA 模板必须被扩增，为下一步的测序反应提供足够的拷贝数。如果按传统方式，将需要几百万个单独的反应管，这是不可能的。微珠被包被在油滴中形成内含微珠的乳胶，由于被包被在自身的反应混合液中，并与邻近的珠子隔离，形成微反应系统。在每个珠子上的 DNA 在各自的反应混合物中 PCR，富集至起始 DNA 模板的 1000 万个拷贝。为完成测序反应，带有 DNA 模板的微珠被放于一个特定的包含 160 万反应池的格子内，这些小尺寸大约 1pl，可确保一个微珠在一个反应池中，通过这样一个格子，超过 100 万的 DNA 模板被扩增。每个模板的 DNA 序列用上述的焦磷酸测序同时进行大量测序。尽管包含单个核酸的反应混合液从格子中被洗下来，

一个灵敏的摄像机能够同时在每个反应池中检测到结合到延伸链的一个核苷酸的发射光。

10.3 基因组测序

第一个被基因组测序的细菌是 φX174（5.4kb）。其他病毒序列包括 λ 噬菌体（49kb），但是测定细菌基因组的全序列（约比 λ 噬菌体大 100 倍），只有通过自动测序仪和机器人联合测定方法才可能完成。

1995 年 7 月 28 日是分子生物学历史的一个重要日期，在这一天，Venter、Smith 和 Fraser 报道了自由存活生物体的第一个完整的基因组序列[嗜血杆菌（*Haemophilus influenzae*），基因组长约 1.83Mb]。今天，通过高通量测序技术从细菌基因组中获得 DNA 序列已成为常规工作。这种测定整个基因组的能力使细菌和整个细菌学领域发生了革命，目前，仅仅在发表了第一个细菌基因组序列的 15 年后，已经有接近 1200 种不同细菌的近 2000 条细菌基因组序列被测定。由于新测序技术的快速发展，这些测序数字将会更大。

这些序列的分析提供了认识生物体的大量生物学信息，也包含了那些使用传统方法不能获得的信息。例如，传统的细菌遗传学很显然只鉴定一些基因组中的一部分基因，有一半左右的预测的编码区过去未经明确定义，并且对功能一无所知。所有微生物中研究最透彻的 *E. coli* 的 30% 以上的基因，仍未知其功能。通过比较整个全基因组序列，可以测定进化压力作用下的细菌生理学。由于显著差异的特性，在同一菌种中很容易选择几种分离子，通过比较它们的基因组序列，可以确定并解释其中一种菌比其他的毒力更强的原因。与之前的遗传学研究不同，在实验室中很容易产生有趣的突变体。到目前为止，即使有基因克隆技术也很难发现基因的功能，而现在通过测定其基因组序列找到变化是很容易的。

基因组计划也是探求意义深远的项目的起点，如鉴定生物体所有基因的功能或定义细菌细胞内所有基因的表达。全基因测序的应用使基因及其功能的大规模研究成为可能，这也被称为基因组学。

10.3.1 基因组测序策略

细菌基因组计划的目标是以低于万分之一的错误率测定数百万碱基的序列。在焦磷酸测序和大规模平行测序之前，当时可以使用的方法使得每个反应最多可以产生 1000bp 的序列，因此要把生物体总 DNA 断裂开，部分测序并组装信息。使用最广泛的策略是全基因组鸟枪法测序，基本的原理如图 10.6 所示，但要使用几百万或几千万个克隆，而不是几个，其原理是一个筛选细菌的 DNA 被打碎成几百万个小片段，并被克隆，建立目标生物体的完整基因组文库，每个插入质粒测序并被组装为重叠群。

由于这些片段为随机片段，其中许多会重叠。因为已被测定的序列为重复测序，这看起来似乎比较浪费。实际上它对进程的成功起着关键作用，不仅因为重叠群测序增加了序列的准确性，而且重叠对于重叠群组装非常关键。

用于新的测序过程同样考虑有两个重要的例外：没有克隆步骤（将使整个进程更快），代替大量克隆工作，所有的片段测序在一次运行中完成。最终获得很多的短片段（比用于脱氧测序产生的片段还短），并最终必须被组装为重叠群。这需要大量的计算机支持。

在基因组测序过程将近完成之后，由于某些区域不能被克隆，经常会有一些缺口存在，如插入序列的重复元件会引起重叠群组装的困难。决定基因组测序速度的限制因素经常靠近这些缺口。具有接受大插入片段能力的载体（如 BAC）在连接大缺口时非常有用，而 PCR 可以连接小的缺口。

与测定最大细菌基因组［天蓝色链霉菌（*Streptomyces coelicolor*）］所使用的测序技术一样，另外一个策略是首先构建有序文库，然后单独测定每个克隆的序列，最后将其连在一起获得完整序列。这种逐步克隆法的优点是每个克隆序列均是相关并有用的信息片段。因为这种方法很少受重复 DNA 的影响，重叠群组装和缺口关闭更容易。然而，建立一个有序文库的工作量很大，一些基因组计划通常联合使用两种方法。

10.3.2 功能相关的序列

生物信息学

细菌的基因组测序仅仅是确定了本身相对无信息的数百万个碱基的顺序。发现这些序列如何决定生物体的生物学是基因组学研究最关键的一步，这要依赖于生物学、计算机科学和信息技术的综合利用，即生物信息学。注释是其中的关键，将基因及其编码蛋白质的功能进行分类，使生物体的全部信息得到组装。在这一过程中，蛋白质的结构和功能特性，包括序列、分子质量、可溶性、活性位点、结合位点、底物、功能、位置、结构、相关通路和相互作用分子将被归类。图 10.9 概括了细菌基因组注释的过程。

基因组注释的第一步是通过计算机程序扫描对可能编码蛋白质或功能 RNA 产物的序列区域进行基因预测。细菌基因组富含基因，80％～90％的序列编码蛋白质，这些程序可以鉴定可读框，转录起始、终止区域和翻译起始位点的保守模式。由于基因序列已被测定，蛋白质序列由此可自动测定并计算其物理特性。

第二步是功能注释。由于在不同细菌中具有相似功能的蛋白质具有相似的结构和氨基酸序列，功能注释的最简单方法是通过与已知蛋白质比较。每一个预测的蛋白质与保存了所有已知 DNA 和蛋白质信息的互联网数据库（有 3 种主要的数据库：GeneBank、EMBL 和 DDBJ，所有的信息在数据库之间可以交换，使用没有问题）中所有的蛋白质进行比较。蛋白质序列的比较比 DNA 序列的比较好，主要是由于生物体间编码区的差异应用，DNA 可以有很大的差异，而蛋白质编码区却相似。与将两个序列排在一起并计算相似率相比，这些序列比较更复杂，幸运的是有程序（使用最广的是 Basic Local Alignment Search Tool，BLAST）可以用于比较（框 10.1）。这个过程表明某蛋白质与数据库中已知或推测功能蛋白质相似，就可以确定这个蛋白质的假定功能。当通过同源性注释蛋白不能发现来源于新测序的基因组序列的蛋白的匹配时，确定功能的第二个工具（序列特点注释）是通过找到蛋白质特定功能特征的短序列，如有疏水性功能区的特殊排列可能是膜蛋白，或表明为 DNA 结合蛋白的预测结构（可能为调节蛋白），有这些特征的计算机

数据库和能与那些文库中的蛋白质进行比较的软件(如 Prosite)均可以帮助确定蛋白质的功能。

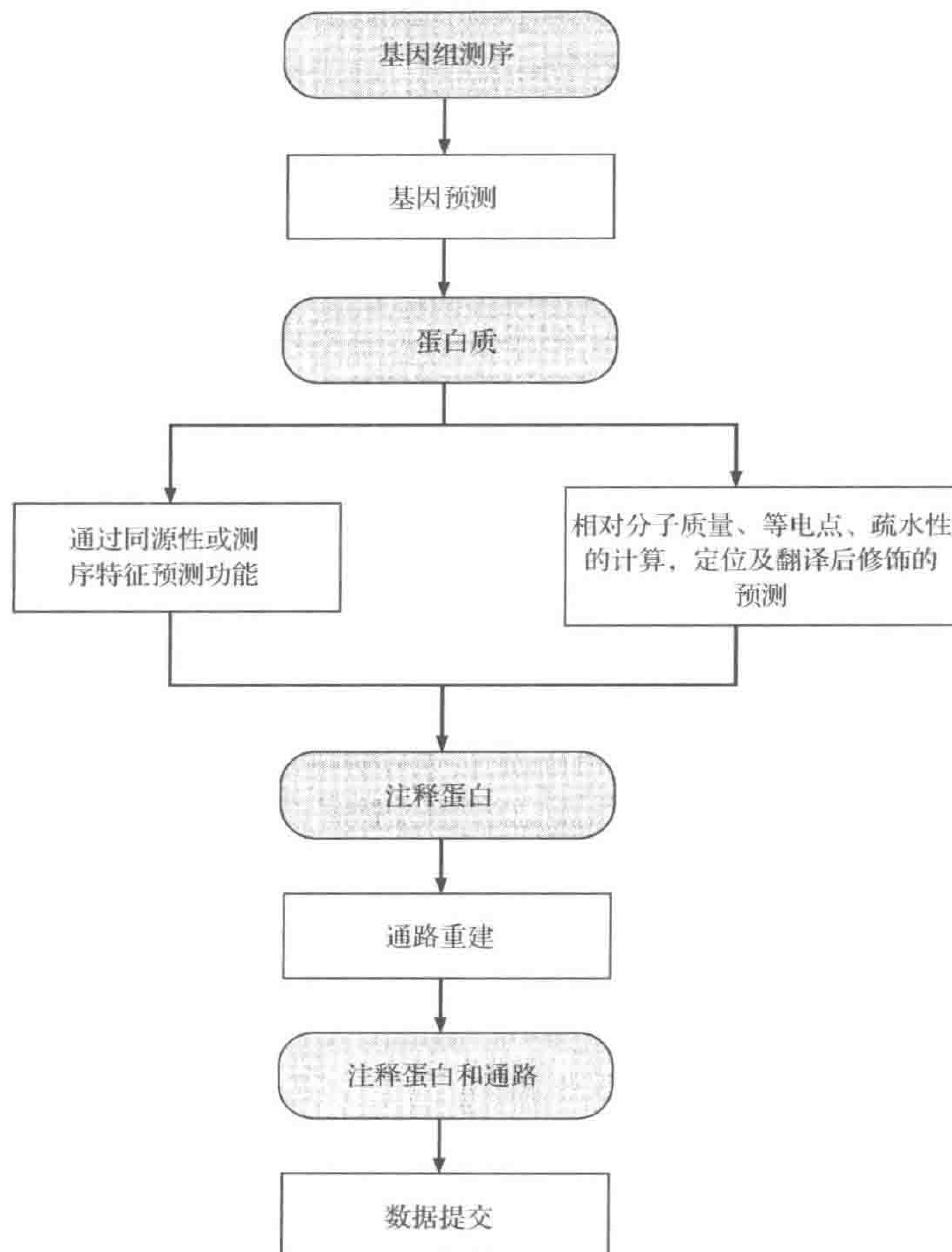


图 10.9 用于注释细菌基因组序列的一般流程

一些蛋白质的单独分离子不发挥作用,但会成为构成许多不同菌种细菌共同拥有的代谢途径的活性通路的一部分。通路重建是在功能注释之上的一种注释过程,可以确定这些构成重要通路的某一部分的蛋白质。有时在细菌间通路是相同的,而有时则是不同的。

注释过程的最后一步是以有效和易于理解的方式建立和显示这些结果,目前是通过建立专用的网站(如 BACMAP)。

基因替代验证基因功能

注释过程仅能提供基因功能的预测,为证实已确定的经预测有功能的基因,通过突变的方法使其功能失活是有必要的。另外,基因的功能与注释过程并不相关,发现其未知功能最简单的办法是突变基因,并使用突变的表型获得对于功能的认识。通过传统的基因

突变和选择技术很难获得失活的基因，特别是如果基因相关表型未知或难以选择。然而，在体外被失活的克隆基因可以通过同源重组的方式，用来替代野生型基因，即基因敲除。整个过程见第 9 章，由于敲除了特定基因，可以观测细胞表型的结果，并确定此基因的功能。

框 10.1　通过同源性及序列比对来举例说明

基因预测程序在最新确定的坂崎克洛诺斯菌(*Cronobacter sakazakii*)(该细菌为阪崎肠杆菌属，能够引起婴儿败血症及脑膜炎，最初是在婴幼儿奶粉中分离得到的)基因组中找到了能够编码下列 211 种氨基酸序列的基因。在其他物理性质中基因预测程序能够自动计算其分子质量为(23 669Da)，氨基酸序列如下。

MEMMMNMSYTLPSLPYAYDALEPHFDKQTMEIHHTKHHQTYVNNANAALE
SLPELANLPVEELITKLDQVPADKKTVLRNNAGGHANHSLFWKGLKKGTTLQ
GDLKAAIERDFGSVEKFKEEFEKAAATRFGSGWAWLVLKGDKLAVVSTANQD
SPLMGEAISGASGYPILGLDVWEHAYYLKFQNRRPDYIKEFWNVVNWDEAAA
RFASQK

为了确定该蛋白质的功能，首先将其与数据库中已知功能的蛋白质进行比较。应用最广泛的是称为局部序列排比检索基本工具(BLAST)的程序，该程序不需要针对全蛋白即可在局部区域发现相似的地方，该方法通过真实序列与人工序列之间的随机性配对判断序列的匹配情况，BLAST 的使用指南可以从 www.ncbi.nlm.nih.gov/Class/BLAST/blast_course.short.html. 获得。

如果将上述氨基酸序列输进 BLAST，将会出现一个序列清单以及 Score、Expect 或者 *E* 值(如下显示)，*E* 值是随机匹配的可能性的指示数值，如果 *E* 值在 e-05 (0.000 01)或者更低，则说明这次匹配是有意义的。Score 为 0.0 说明蛋白质之间是相同的或者非常接近。

```
>ref|YP_001572545.1|  superoxide dismutase [Salmonella enterica subsp. arizonae
serovar 62:z4,z23:--]
  gb|ABX23403.1|  hypothetical protein SARI_03588 [Salmonella enterica subsp.
arizonae serovar 62:z4,z23:--]
Length=210

 GENE ID: 5764305 SARI_03588 | superoxide dismutase
[Salmonella enterica subsp. arizonae serovar 62:z4,z23:--]

 Score = 421 bits (1083), Expect = 2e-116, Method: Compositional matrix adjust.
 Identities = 202/211 (95%), Positives = 208/211 (98%), Gaps = 1/211 (0%)

Query  1    MEMMNMSYTLPSLPYAYDALEPHFDKQTMEIHHTKHHQTYVNNANAALESLPELANLPVE  60
            MEM+ MSYTLPSLPYAYDALEPHFDKQTMEIHHTKHHQTYVNNANAALESLPE ANLPVE
Sbjct  1    MEMI-MSYTLPSLPYAYDALEPHFDKQTMEIHHTKHHQTYVNNANAALESLPEFANLPVE  59
Query  61   ELITKLDQVPADKKTVLRNNAGGHANHSLFWKGLKKGTTLQGDLKAAIERDFGSVEKFKE  120
            ELITKLDQVPADKKTVLRNNAGGHANHSLFWKGLKKGTTLQGDLKAAIERDFGSVEKFKE
Sbjct  60   ELITKLDQVPADKKTVLRNNAGGHANHSLFWKGLKKGTTLQGDLKAAIERDFGSVEKFKE  119
```

```
Query  121  EFEKAAATRFGSGWAWLVLKGDKLAVVSTANQDSPLMGEAISGASGYPILGLDVWEHAYY  180
            EFEKAAA+RFGSGWAWLVLKGDKL+VVSTANQDSPLMGE ISGASG+PILGLDVWEHAYY
Sbjct  120  EFEKAAASRFGSGWAWLVLKGDKLSVVSTANQDSPLMGEVISGASGFPILGLDVWEHAYY  179

Query  181  LKFQNRRPDYIKEFWNVVNWDEAAARFASQK  211
            LKFQNRRPDYIKEFWNVVNWDEAAARFA++K
Sbjct  180  LKFQNRRPDYIKEFWNVVNWDEAAARFAAKK  210

gb|AAT99791.1| manganese-dependent superoxide dismutase [Pasteurella stomatis]
gb|AAT99792.1| manganese-dependent superoxide dismutase [Pasteurella stomatis]
Length=150

Score =  237 bits (605),  Expect = 5e-61, Method: Compositional matrix adjust.
Identities = 116/150 (77%), Positives = 127/150 (84%), Gaps = 1/150 (0%)

Query  23   HFDKQTMEIHHTKHHQTYVNNANAALESLPELANLPVEELITKLDQVPADKKTVLRNNAG  82
            HFD  TMEIHHTKHHQ YVNNANAALE+LPE A+    EL++KL  VPADK T LRNNAG
Sbjct 1     HFDAMTMEIHHTKHHQAYVNNANAALENLPEFASACPGELLSKLSLVPADKFTALRNNAG  60

Query 83  GHANHSLFWKGLKKGTTLQGDLKAAIERDFGSVEKFKEEFEKAAATRFGSGWAWLVLKGD  142
            GH NH+LFWK LKKGTTLQG LK AIERDFGSVE F+ EFEKAAATRFGSGWAWLVL+ D
Sbjct 61  GHLNHTLFWKSLKKGTTLQGALKDAIERDFGSVEAFQAEFEKAAATRFGSGWAWLVLQAD  120

Query 143 -KLAVVSTANQDSPLMGEAISGASGYPILG  171
             KLAVVSTANQDSP+MG+ I+G  GYP+LG
Sbjct 121 GKLAVVSTANQDSPVMGKEIAGCEGYPLLG  150
```

序列清单中可能包括上百个相似的结构,这里只列出了两个典型的例子。上层标记为 Query 的序列是来自 *Cronobacter sakazakii*,下层的 Sbjct 序列是经 BLAST 鉴定后与 Query 序列相似的序列,中间序列是 Query 与 Sbjct 序列相同的部分。在两组蛋白质序列中,如果某点氨基酸相同,则会出现相应氨基酸的代码,如果两组序列中氨基酸序列不同则会出现空格。缺失则用破折号代替。目测可知,未知蛋白质与已经确定含有过氧化物歧化酶功能的蛋白质具有很高的相似性。因此,经序列比对可知道该蛋白质在坂崎克洛诺斯氏菌中同样具有过氧化物歧化酶功能(也可说明它们具有同源性),该蛋白质也可以被假定为过氧化物歧化酶(之所以假定是因为该蛋白质的活性还需要实验证实)。E 值越低或者越接近 0,匹配越具有意义,通过与肠道沙门氏菌分枝杆菌(*Salmonella enterica* subsp. *arizonae*)和巴氏杆菌(*Pasteurella stomatis*)的相似性进行比对,其 *E* 值分别为 2e-116 和 5e-61,因此可以判断其与肠道沙门氏菌分枝杆菌比较更有意义,然而,当这二者的 E 值均小于 e-05 时,二者都有意义。

基因功能的系统分析

基因组序列就是一个细菌全部基因的目录,而其中仍有很多被确定的基因的功能未知。因此,在某些细菌全基因测序完成之后,分析所有已知阅读框基因的功能的系统尝试已经开始实施。其目标是针对个别菌种失活每个基因,通过分析表型结果来确定每一个蛋白质的功能。就现在而言,这些宏伟的目标将只有在获得细菌基因组完整序列的前提下才可能得到实现。

定点突变

相对于将一个基因完全敲除，用一个特异改造过的基因将其取代更能满足人类的愿望。为了达到这样的目的，应该将前面描述的基因置换和定点突变技术结合起来，如第 8 章讲述的那样，可以通过直接置换，也可以利用携带有突变基因的质粒补充敲除掉的基因。

例如，利用一系列不同的氨基酸代替某个氨基酸并检测终产物的活性，可以鉴定特定氨基酸在某个酶中的作用，或者通过把一个明确了突变位点的序列引入启动子区域，可以研究调节序列发挥作用时对于 DNA 的精确需求是什么。

10.3.3 宏基因组学

传统的细菌基因组测序工程始于实验室同样的细菌培养物及其为了用于特定目标的 DNA 资源。对于能够在实验室培养的菌株，这种方法非常有效，但却难以适用于那些不能在实验室培养的菌株。在过去的 15 年中，所得到的细菌基因组序列让这个问题显得不那么严重，但是随着对生长在自然环境中的细菌的了解越来越多，这个问题又变得棘手起来。据当前估计，实验室培养得到的细菌低于环境中所有细菌的 1%，因此，所掌握的大部分关于细菌的知识仅来源于能够在实验室培养的很少的那部分细菌。

宏基因组学是一门新兴技术，该技术不需要分离和培养细菌，可以直接利用从自然环境中提取的 DNA 的遗传信息来研究细菌，这就解决了一些细菌在实验室不能培养的问题。该技术是在 20 世纪 80 年代研究 16S rRNA 时出现的。这些序列及它们的种系发生在第 9 章中均有详细的介绍，这里要重点说明的是该序列可以作为每个细菌品种的遗传条码，可以用来鉴定一些不经过培养的菌株。在早期研究中，利用 PCR 方法从存在于所选择的环境样本中的细菌 DNA 中扩增 16S rRNA。单个序列被克隆并测序。这种方法曾经适用于很多环境，后来发现，存在于环境样本中的很多 16S rRNA 序列不属于任何已知的被培养的菌株。现在，除了 16S rRNA，环境中很多 DNA 序列也经常会用到，该方法也可以用来识别尚未被培养过的工业中重要的菌源性工具酶。因为通过该技术，科学家可以接触到由于无法培养之前没有研究过的数百万计的微生物(更重要的是其全新的遗传信息)，宏基因组学正在使微生物学发生革命。

一个典型的宏基因组学项目是从混合菌群 DNA 的分离开始的，菌群可以从任何一种环境(如牙菌斑、动物内脏、极端环境或者人的手上)收集而来。得到的 DNA 经片段化后用来组建 DNA 文库，可用于随机或靶向测序。随机测序和鸟枪法测序相类似，克隆被随机选择，序列结果通过重叠序列的匹配组合到更大的连续组合中。在靶向测序中，克隆首先根据合适的基因进行筛选，如利用杂交或 PCR 技术，也可以通过基因功能(如酶分析)进行筛选，只有具有合适基因的序列才会被筛选并用于序列分析。在以上两种情况下，得到的 DNA 序列会与数据库进行比对并注释，如果可能的话，将采用生物信息学对以上情况作出描述。

随机测序提供了发现完全未知基因的潜力，当这种技术和上述描述的一些测序技术相结合，就会省略掉基因克隆的步骤。该技术创造了一种非常强大的对任何环境中的菌株差异性进行鉴定的方法。有两个例子可以进一步说明其强大的能力，2006 年，利用该

技术检测出了生长在矿井酸水池中生物膜的全基因序列 76.2Mb，包括两个先前没有被培养过的生物体的近似完整的基因组序列，部分序列是三种生物体所共有的。另一项研究采用马尾海藻样本产生的高于 1.6Gb 的序列，发现了先前并不知道的 120 万个未知基因序列。

10.4　比较基因组学

随着越来越多的细菌的基因组序列的破译，对它们进行比对并验证基因功能以及获得细菌的进化历史的信息成为可能。例如，如果对比两个菌株中的 RNA 聚合酶，则会发现它们具有非常相似的基因序列。它们是源自共同的原始分子，这两种酶称为同系物。很明显，在这个例子当中，这两种同源酶是由相同的原始基因经过特殊方式演变得到的，它们具有相似的功能，因此这两种酶也可以称为直系同源(同源酶的特殊形式)。有时进化的过程并不是直接发生的。例如，旁系同源物是由基因组中某基因的副本产生的。因此，虽然它们共用一些氨基酸，但具有不同的功能。相反，基因或者蛋白的类似物虽然由不同的原始分子演变而来，它们之间没有任何的关系，但是却具有相同的功能，因此它们被称为非同系物。

基因组序列中的一个最大区别是基因组序列大小不同。黑草属(*Buchnera*)和生殖支原体(*Mycoplasma genitalium*)的基因组大小分别为 450kb 和 580kb，铜绿假单胞菌(*Pseudomonas aeruginosa*)的基因组大小为 6.3Mb，百脉根根瘤菌(*Mesorhizobium loti*)的基因组为 7.04Mb，链霉菌(*Streptomyces coelicolor*)为 8.7 Mb，纤维素堆囊菌(*Sorangium cellulosum*)为 13.0Mb。黑草属是一种蚜虫细胞内共生生物，进化后只保留了仅能够维持其在自身环境下生长的最少数量的必须基因。另外，铜绿假单胞菌是一种适应能力非常强的细菌，能够在包括人体在内的很多环境中生存。天蓝色链霉菌(*S. coelicolor*)是一种在土壤中生长的细菌，其生长周期包括不同形态的阶段(详见第 9 章)，也具有营养多能性和制造次级代谢物的能力(链霉菌属是自然抗生素的主要来源)。毫不奇怪的是，这两个生物体都含有大量的调节基因，并且其基因组的大小和复杂性反映了细菌能够适应不同环境的进化适应性。

基因组首先被破译的两个细菌是流感嗜血杆菌(*H. influenzae*)和生殖支原体，每一个基因序列很快就应用到对于其生物学特性的认识上。例如，由于流感嗜血杆菌需要大量的谷氨酸盐，在实验室中并不常见，其基因组中缺少编码谷氨酸盐的密码子，因此，谷氨酸盐能够弥补这种不足。在生殖支原体中，其基因组中完全没有参与氨基酸合成、核苷酸合成、脂肪酸合成的酶的基因，这就解释了这种细菌的生长对环境的要求很严的原因。

研究大肠杆菌比研究任何一种其他微生物所获得的生物学认识都要多。大多数大肠杆菌菌株能在内脏中共生，只有一小部分有致病能力。因此，非致病性菌的基因序列为研究其他大肠杆菌的基因序列提供了一个基础。例如，大肠杆菌 O157：H7 是威胁食品安全的最危险的病原菌之一，对其基因组测序发现有超过 1000 个基因是非致病菌株 K-12 所缺乏的。这些所谓的新基因似乎与 O157：H7 的致病能力有关。这些插入 DNA 的性质及起源在第 4 章中有详细的介绍。

耶尔森氏鼠疫杆菌(*Yersinia pestis*)的演变是高致病性病原菌出现的一个典型的例子。该细菌从假结核耶尔森氏菌(*Yersinia pseudotuberculosis*)进化而来,可以引起人体内脏轻微的疾病,进化了 2000～20 000 年,这些细菌可以引起广泛的疾病,但它们非常接近(在 DNA 水平有 97%是一致的),现在通过基因组比较,知道了一种细菌如何从另一个种细菌演变而来。现在,许多细菌的基因组已经确定,其中一部分是不同来源 DNA 序列的拼接,由于大的可移动遗传元件的交换,如质粒或毒力岛(详见第 5 章和第 9 章)。这个过程可以使细菌病原体以所谓的量子跃进的方式发生快速进化。在耶尔森氏鼠疫杆菌中,获得 DNA 质粒是导致其从温和型向高致病性病原体状态转变的关键因素。其中 pPla和 pMT1 两种质粒是该菌种特有的,由于为通过蚤传播瘟疫杆菌所必需的,使其获得在改变耶尔森氏鼠疫杆菌的传播周期中发挥了很重要的作用。相反,当微生物适应新的环境时,经常会出现基因功能丧失或者基因组衰减。假基因,与已知基因相关的不起作用的基因,保持着它们的同源性,但由于突变的积累导致其不能编码蛋白,在病原体为了适应新环境压力而减少其基因组量时经常会出现。这类基因占到耶尔森氏鼠疫杆菌基因组的 4%,很显然,当病原体改变了其生长状态,从一个感染了的个体的内脏中经过蚤传播到另一个个体并感染,许多基因对于其在已经腐烂的内脏中生长是必需的。

引起炭疽的炭疽杆菌(*Bacillus anthracis*)在环境中会形成静止的芽孢,使其不受环境的影响,只有在体内被巨噬细胞识别后才会引起疾病。众所周知,该菌是一种专性高致病性菌,在体外不能生长,但是该菌是从生长在土壤中的原始菌进化而来的。这种生活方式的戏剧性的改变,给炭疽杆菌及其基因组带来了很大的选择压力,如同耶尔森氏鼠疫杆菌,这种变化的线索存在于其基因组序列中。炭疽杆菌具有两个大的致病质粒,pXO1 和 pXO2(第 5 章所示)。这些质粒是非致病菌所不具备的,获得这些质粒可能是炭疽杆菌进化为致病菌过程中的关键步骤。另外,当把致病菌与生长在土壤中的菌进行比较时发现,致病菌的基因组中没有编码降解多聚糖的酶的许多基因,这对于其在过去利用土壤中的碳资源非常有用,但是在感染的过程中没有作用,作为基因组衰减的结果,该部分基因就丢失了。

炭疽杆菌、蜡样芽孢杆菌(*Bacillus cereus*)(一种引起温和的食物传播的疾病的病原菌)和苏云金杆菌 (*Bacillus thuringiensis*)(可引起昆虫疾病但不引起人的疾病)在遗传水平上非常相近,事实上曾经被认为是同一菌种。对于鼠疫杆菌和假结核耶氏菌而言,其基因组在菌株间的差异很小,而质粒很大程度决定了在毒力上的巨大差异。在其他细菌中,基因组比较揭示了一系列不同的现象,在这些例子中,在同一菌种基因组间的广泛遗传变异要比想象的多得多。多达 20 种甚至以上的大肠杆菌基因组测序证实了这一点,同一菌种的基因组大小为 4.6～5.5Mbp,菌株基因组间的变化高达 25%。大多数的突变源于噬菌体。对于这些细菌,至少传统的教科书观点认为进化主要基于相对缓慢的个体散在的突变的累积而使基因组改变。而这里难以想象的菌种间高度差异表明,来源于一个菌株的单个基因组序列并不真正代表一个品种的详细目录,由于它仅代表一个样品的基因,而广泛的基因池基因代表着品种。这就衍生了泛基因组的概念,在每个品种中均存在着在所有菌株中都保守的、包括一部分共同拥有的基因或菌株特异性的特有基因的一整套核心的基因。

10.4.1 微阵列

基因组比较最简单、直接的方法是比较序列，但成本限制了被测序的基因组的数量，尽管这种现状正有所改变。另外一种方法是使用微阵列技术。通过微阵列技术可以研究应用一种细菌的基因组序列研究在同一菌种中其他菌株出现的变异，如一旦获得了结核分枝杆菌强毒株的基因组测序结果，就可能找到卡介苗(bacille calmette-guerin，BCG)与已测序的结核分枝杆菌强毒株的不同(由于 BCG 的基因组序列已经被测通，这个例子只具有历史意义)。微阵列的进一步的应用即比较两个生物体的基因表达，将在本章的后面部分详细介绍。

基于细菌的全基因组序列(如结核分枝杆菌)，利用 PCR 扩增或使用寡核苷酸合成技术，产生一系列的与基因组中每个基因相关的 DNA 片段是可能的。这看起来好像是不可能完成的任务，如结核分枝杆菌有大约 4000 个这样的 DNA 片段，如果手工做的话确实很艰巨。事实上，通过机器人设备操作，可以在类似网格排列玻璃板表面"印"上每个寡核苷酸的微斑点(再使用一个机器人给出每个斑点的准确位置)。

为了比较两种细菌(结核分枝杆菌和 BCG)的遗传信息，从每个生物体中提取 DNA 并标记不同的荧光染料。然后两个 DNA 样本混合并加到玻璃片上，使其与微阵列的斑点杂交(图 10.10)。在微阵列上，标记片段与斑点结合，可通过设备区分两种不同的荧光

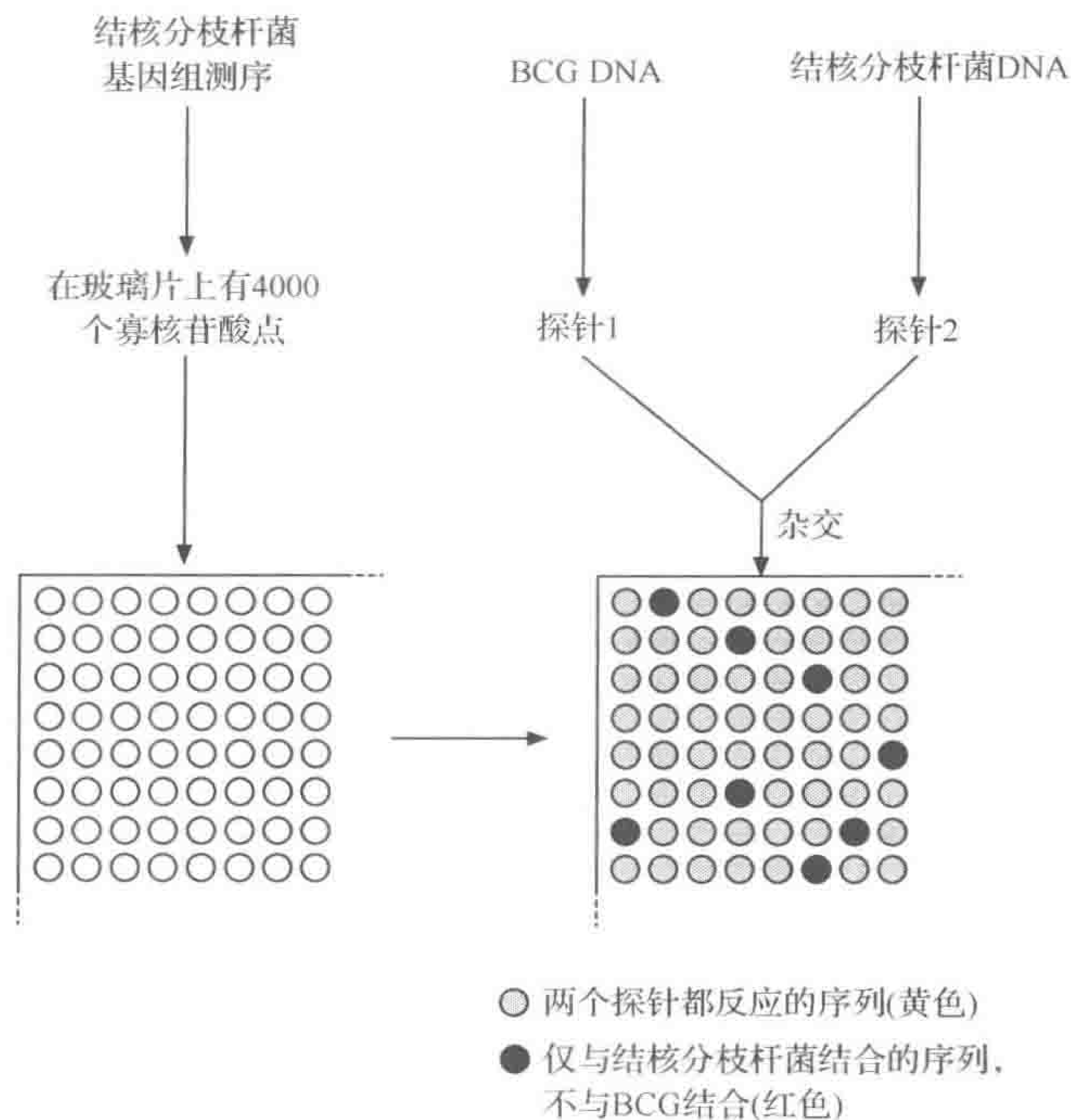


图 10.10 使用微阵列比较基因组

两个探针标记了可被微阵列机器自动识别读取的不同荧光染料。与探针 2 反应的杂交点显示为红色，而与两个探针均反应的显示为黄色，仅与探针 1 反应的显示绿色，在本例中，不会预期探针 2 会与同样来源样本的所有点杂交

染料，结果通过点图谱显示出来，红色或绿色代表仅与一种或另一种样本结合，而黄色代表两种标记DNA均结合。因此，在本例中，结核分枝杆菌和BCG的共有基因表现为黄色，而在结核分枝杆菌中仅有的基因为红色。一个重要发现是DNA区域(如RD1)在BCG不存在，这一区域包含与毒力相关的数种基因。微阵列并不能告知BCG是否有不存在于结核分枝杆菌中的其他DNA，因此，不能对任何绿色的圆点有期望。

10.5 基因表达分析

当面对一个新环境时，细菌可识别许多不同的环境刺激并改变它们的基因表达方式来优化生存(见第3章)。伴随着细胞在形态或生理阶段变化，或在不同外部环境中(如在巨噬细胞中细菌的生长)，通过追踪基因表达的改变(在孢子形成过程中或过渡到稳定期)，了解很多基因的功能，从而可以提高对于这些生理反应的认识。

通过分析酶活性的最常见的监测基因表达的方法不总是合适的。首先，并不是所有的基因产物都是酶，一些最有趣的基因产物是调节蛋白。其次，即使是酶，分析过程也可能太复杂而不能用于此目的，但可以通过下面所述的技术评价基因表达。

10.5.1 转录分析

在细菌中，基因表达主要是在转录水平进行调节，因此能够合理地推测测定细胞内特定mRNA的数量是测定基因表达的有效途径(尽管情况并非总是如此)。总体而言，细菌mRNA寿命短，所以在一定的时间内测定一种特定mRNA的量是评价该基因当时的转录水平的一种好方法。在第9章中，已提到了报道基因的使用，这仅是在体内利用的方法，在这里关注的是在体外用于测定mRNA水平的方法。

Northern 印迹法和反转录 PCR

确定特定mRNA产量的一种方法是通过凝胶电泳对细菌RNA进行分级，并通过印迹将RNA转印到膜上，然后将膜与适当的探针杂交。信号的强度提供一种特异mRNA定量的方法。由于这一过程与Southern印迹法相似，因此被称为Northern印迹法。

Northern印迹法并不是理想的方法，首先，一些细菌mRNA非常不稳定，半衰期为1～2min，其次，在检测的条件下难以获得足够多的RNA。例如，检测在单个克隆的不同部分的表达水平，或者在巨噬细胞内细菌的生长。一种更敏感的技术是PCR的采用。

在标准的PCR(见第2章)中，使用DNA聚合酶来制备模板DNA的拷贝。为了获得检测特定mRNA的灵敏方法，使用RNA介导的DNA聚合酶，即反转录酶来产生一个DNA拷贝，反转录后使用标准的PCR对所产生的DNA链进行扩增。这种反转录PCR或RT-PCR方法(图10.11)，为在菌群中的小部分特定转录本的检测提供了一种非常灵敏的方法。

实时定量 PCR

如上所述，RT-PCR的一个最大的局限为不容易定量。只能用于定性检测一个转录本，但是不能定量地反映这个转录本的数量。对此转向另一种变化的PCR技术，实时定

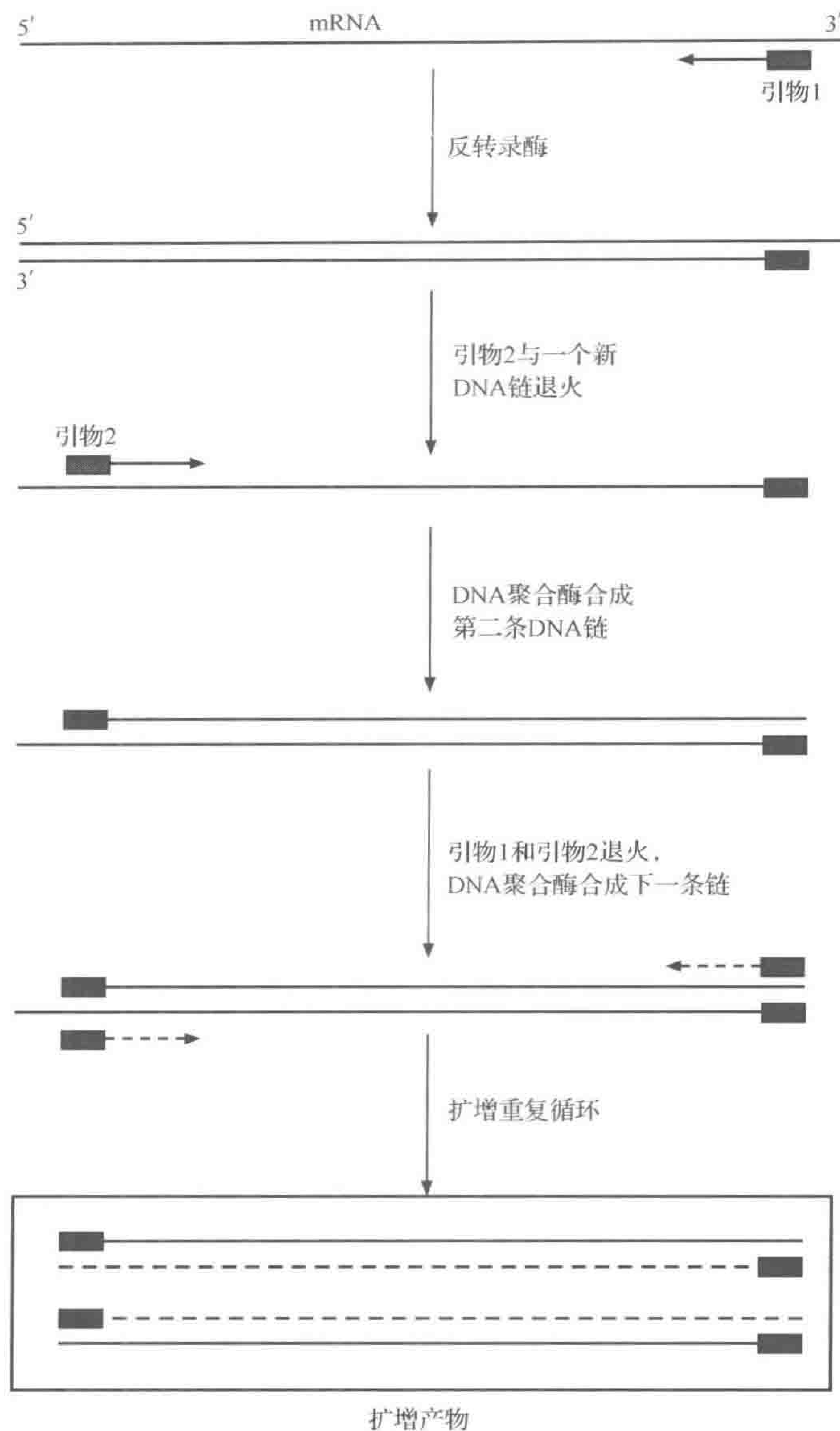


图 10.11 反转录 PCR

量 PCR 技术，因为在 PCR 反应期间它能持续地监测扩增产物的出现。

实际上有几种不同的技术可以达到这一目的，但是实时定量 PCR 是最简单的。这涉及在 PCR 反应混合物中的一种荧光染料，它可以结合到双链 DNA 分子上并发出荧光。通过机器可以检测到这些信号，该机器不仅能够完成温度循环，而且还可以测定样品的荧光信号。最初 DNA 的浓度非常低，因此没有信号被检测到。但是随着反应的进行，通过一个循环到另一循环，扩增产物的量将以指数方式增加，直到 PCR 仪上显示这一信号(图 10.12)。需要达到这一临界点的循环数依赖于初始模板靶 DNA 的量。因此，图中样品 A 仅需要 20 个循环就能达到检测水平，与样品 B(22 个循环)或者样品 C(25 个循环)

相比，样品 A 包含了更多的模板量。

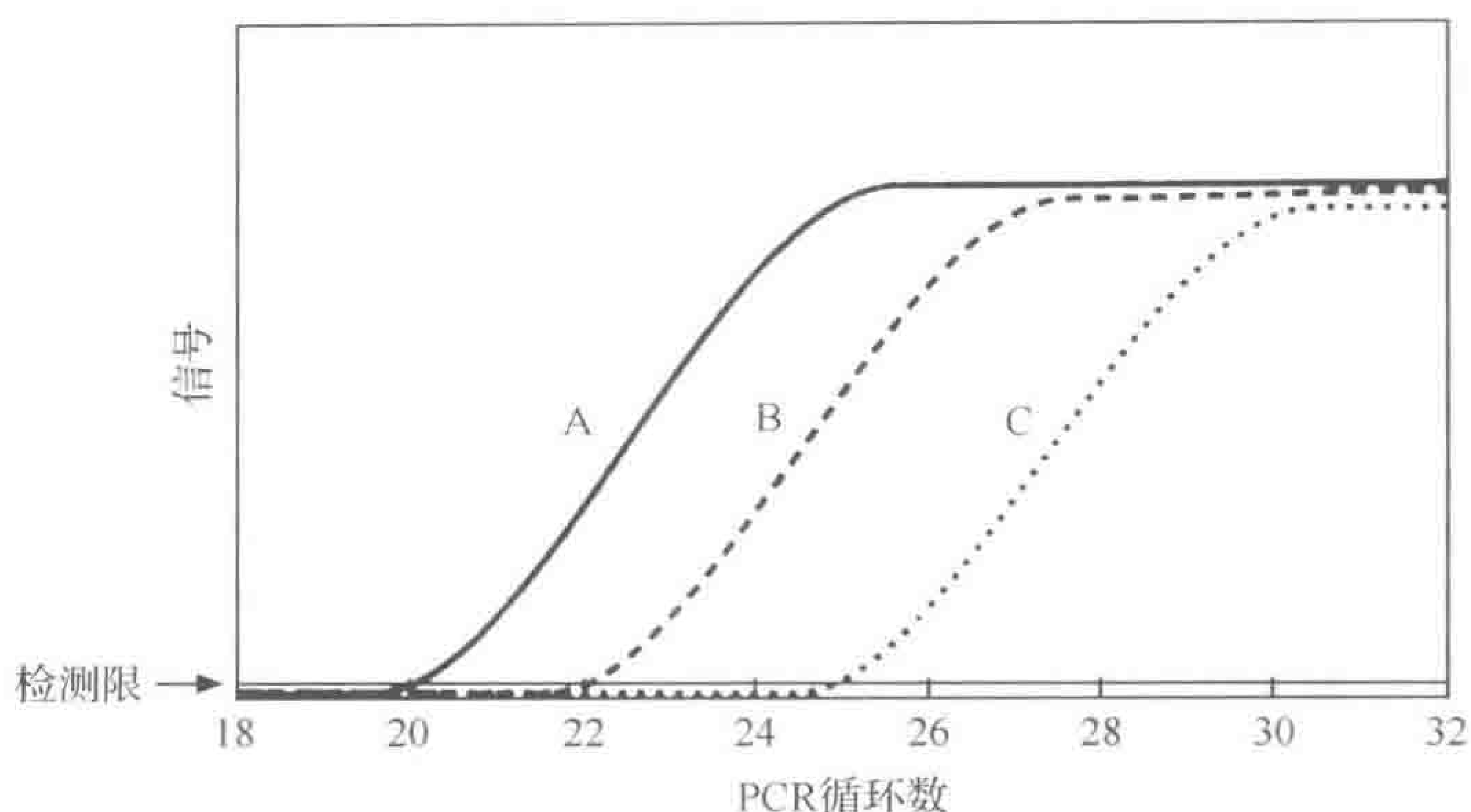


图 10.12 实时定量 PCR

随着 PCR 进程可以检测到数种产物。在检测信号前需测定循环数，最终的模板量可被估算

当然，对于一个转录本的检测水平，开始于 RNA 而不是 DNA。因此，在实时定量 PCR 前进行反转录是必需的。

实时定量 PCR 的优点除了定量以外，还在于产物检测速度和容易操作，不需要凝胶电泳。这个在 PCR 的诊断应用上是非常有用的(第 9 章)。另外，在基因表达变化中已经应用微阵列技术去鉴定，这个将在下面的内容中描述，实时定量 PCR 也通常独立地应用于证实基因表达的变化。

应用微阵列技术分析 mRNA 的表达

前面所介绍的分析基因表达的方法，以及报道基因的使用被设计用于单独的少量的基因表达(第 9 章)。然而，随着全基因组序列的应用，现在在细菌中可以应用微阵列技术同时评估所有基因的表达。通过基因组大规模地研究 mRNA(或者其他类型的 RNA)的表达被称为转录组学(transcriptomics)。

在本章的前部分已介绍了应用 DNA 样本进行微阵列的概念。为了评估基因表达或者每个 mRNA 的特异表达量，要将标记的 mRNA 替代标记的 DNA(mRNA 通过反转录产生标记的 cDNA)。

为了评价微阵列上的基因表达，在对照条件和兴趣刺激条件下细胞生长(图 10.13)，如比较活跃生长细胞(培养物 A)和稳定期细胞(培养物 B)的基因表达。从两个培养物中提取 mRNA，反转录产生一个 cDNA 分子谱。每种 cDNA 被单独标记不同种荧光染料，cDNA 样本混合并孵育到微阵列上，并与微阵列的斑点杂交。带有不同染料的 cDNA 混合样本与每个微阵列的斑点竞争结合，结果依赖两个样品中特异 cDNA 的相对量。之后，微阵列将应用如前所述的特定阅读器扫描。在这个例子中，旺盛生长期间表达量高的基因多来自于带第一种染料标记的 cDNA，来自染料的信号将更强，在图中显示的是红斑点(spot 1)。相反，在稳定期基因被激活后，cDNA 标记第二种染料，将显示绿色的斑点(spot 3)。如果一个基因在两个条件下同时表达，信号将在两种染料之间没有差别，将显示出黄色斑点(spot 2)。

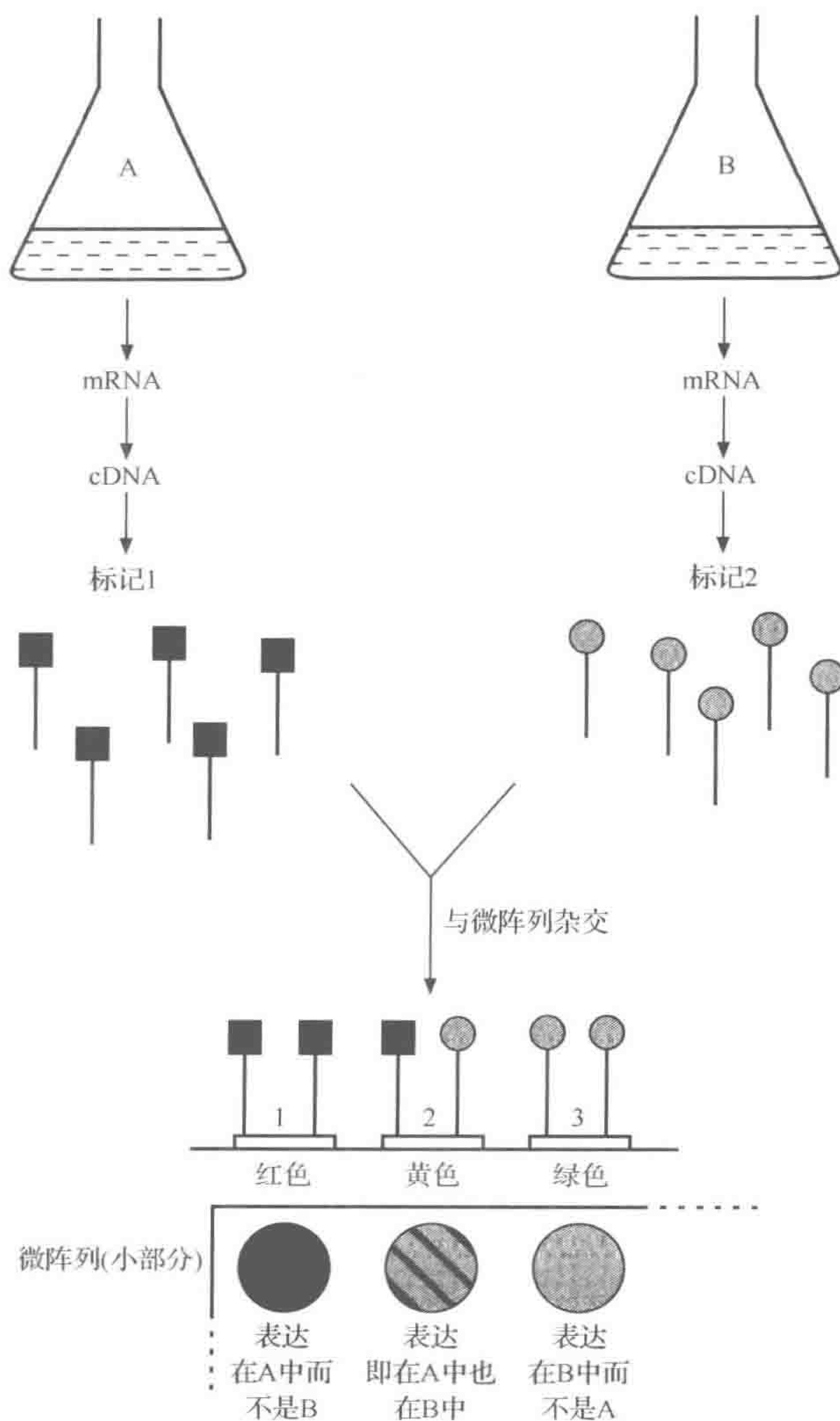


图 10.13 使用微阵列分析基因的表达

来自于培养物 A 和 B 的 cDNA 探针标记不同的荧光染料，如图 10.8 所述，如果某个基因在培养物 A 中表达较多，标记 1 将与微阵列中的点结合更紧密，表现为红色。在培养物 B 中某个基因表达更强，则表现为绿色，若没有差别则表现为黄色

因此，在这些条件下，通过简单斑点颜色，能评估基因组中所有基因的相对表达量。在一种特殊的条件下，一个细胞的基因表达谱将被转录，被称为转录组学，因与基因组学相似而得名的，包括了一个细胞的全部遗传内容。值得注意的是，基因组对于单个的菌株或多或少是固定的，而转录组不同，将会由于收获细胞的时间不同而产生不同的结果。

转录组测序

通过分子标记并在预先设定位置的网格上结合到它的同类 DNA 分子，微阵列可以检测一个特殊的 mRNA 分子的表达。随着大规模平行测序的应用，转录组测序不久可能

绕开微阵列技术，测序即可鉴别一个细菌细胞产物的所有类型的不同 mRNA。对于这种方法的深层讨论不在这本书的范围，在这儿仅仅是一个简短的略述。在选定条件下，一个生长细菌的 mRNA 总量能通过反转录转换成 cDNA 文库，之后应用新的测序技术能很快地测序。这些序列信息可用于鉴别 mRNA（和它的同类基因）和它的相对表达量（mRNA的含量越高，它在文库中出现的次数将越高）。这个方法在不久的将来将替代微阵列技术。

10.5.2 翻译分析

正如前面提到的一样，mRNA 表达水平的分析是基于假设基因调控发生在转录水平，但是这可能是不对的。应该考虑到这种可能性，特殊蛋白的表达水平甚至存在会受到翻译效率、其他形式翻译调节的影响。如果有一些相似基因编码不同的蛋白质，由于转录产物相互杂交，通过转录的分析可能很难区分它们。

Western 印迹法

如果有可识别蛋白的特异性抗体，就可以利用它们检测不同生长条件下蛋白的表达。基于此，通过聚丙烯酰胺凝胶电泳（SDS-PAGE）分离细胞提取物，将蛋白转移到膜上，并使用抗体检测特异性蛋白。这些抗体可以直接标记或者使用标记的二抗检测。大部分标记物是酶（如过氧化物酶），它能通过与显色底物反应而被看到。这个实验过程被称为 Western 印迹法，尽管这个过程相当的不同，要涉及电转移，但与第 2 章介绍的 Southern 印迹法相似。然而 Western 印迹法仅适合于检测特异性蛋白质的表达。对于蛋白质表达的全部分析，要求诸于蛋白质组学研究。

蛋白质组学

本章至此，已经涉及了基因组和 RNA 的大规模研究——基因组学和转录组学，随着高通量技术的发展它们已成为可能。同样的，蛋白质组学是一门大规模研究蛋白的学科，这将与那些其他的“组学”技术补充和重叠。蛋白质组学的总体目标是对于给定生物体的全部蛋白质（又称为蛋白质组）进行系统的鉴定。与基因组学相比，蛋白质组学更具有挑战性，这是因为一个细菌的基因组是相对恒定的，而蛋白质的表达则具有时间性和空间性。

蛋白质组学的另一目标是鉴别蛋白质的功能和机体范围内的相互作用，而这一目标的实现，只有在基因组测序和生物信息学的发展前提下才会成为可能。许多蛋白质在翻译后要经过磷酸盐、脂类、碳水化合物等的化学修饰，例如，蛋白质组学另一方面是在整个机体水平而不是单独寻求已经修饰的特性蛋白。通过使用生物信息学的计算机程序可以预测修饰的靶蛋白的保守氨基酸序列，也可以通过制备抗体结合或关注修饰的蛋白。

应用蛋白质组学分析基因表达

应用微阵列的方法可以在整体水平上分析蛋白质表达水平。但是，由于 mRNA 不总是翻译成蛋白质，从已知的 mRNA 产生的蛋白质的量是不同的，因此，微阵列分析结果也不总是精确地反映蛋白质的表达量。因此，由于能够直接测定细胞内目标蛋白的表达

水平，蛋白质组学提供了一个更可靠的蛋白表达分析方法。用于蛋白质组学的技术是特定的，并且大部分超出了这本书的范围，但是它包含了分离和可视蛋白的技术，鉴定蛋白质-蛋白质相互作用的技术，鉴定翻译后修饰技术等。

细菌细胞可能包含几千种不同的蛋白质，因此，需要有效的方法来分离它们。蛋白质的氨基酸序列决定了蛋白质的分子质量和电荷总数，通过这两个特性可以分离细胞中的单个蛋白。例如，许多蛋白质有着相似的分子质量，不能通过 SDS-PAGE 分离。但是这些具有相似分子质量的大部分蛋白质的酸性、碱性氨基酸组成不同，这将决定了其不同的总电荷数。蛋白质的电荷特性依赖于 pH。在高 pH 时，它们带负电，在低 pH 时，带正电。两个值之间为等电点(isoelectric point，pI)，蛋白质呈中性。不同蛋白质的等电点因氨基酸组成的不同而不同。

2D 聚丙烯酰胺凝胶电泳(two-dimensional polyacrylamide electrophoresis，2D-PAGE)分离蛋白质依赖于连续的两步：电荷和分子质量(图 10.14)。第一步与等点聚焦(isoelectric focusing，IEF)相关，在细胞提取物中蛋白质分散在两性电解质的胶中，通过使用化学方法获得稳定 pH 梯度。在胶中加一个电场，每种蛋白质通过梯度迁移到达等电点，净电荷成为零而静止不动。由于等电点不同，一些蛋白质将终止于 IEF 胶的不同位置。然而，由于一些蛋白质的等电点相似，情况并非完全如此。在 SDS-PAGE 中应用 IEF 胶可以只依据其分子质量分离蛋白质。使用此项技术，可以同步分离 3000 余种蛋白质，因此可以在一些选择条件下研究细菌细胞中几乎所有蛋白质的表达谱。

跑胶后，将胶染色，可溶性蛋白将以不同的分离点形式出现。最终，已知蛋白质在胶中的位置表明其由电荷和分子质量所形成的独特标记，可用于鉴定。另外，通过从胶中切下蛋白点进行微化学分析，可以鉴定一种未知蛋白质。在过去，蛋白质的氨基酸测序已经用于其鉴定，因为能提供一部分但特异的氨基酸序列，通过反向翻译能追溯到基因组中相关的基因。目前这项技术已被质谱仪(MS)的使用大大提升。

许多类型的质谱仪及其深入介绍并不是本书讨论的范围，但均为同样的通用技术。被鉴定的蛋白质被注入 MS 的真空室内，并暴露于电子源。因为 MS 仅能作用于带电的分子，这些电子碰撞掉肽的其他电子并使之带正电，这个过程很重要。然后，离子化的肽被加速，并以同样的高速注入远端带有检测器的磁场。一旦进入磁场，离子的转向与它们的质量(越轻越容易转向)和离子的正电荷数量(交换的离子越多，越易转向)有关，每种离子类型在磁场中将有所不同。磁场的性质可以被改变，因此，在特定的磁场强度下，每种离子可被特异地传送到检测器，由于这些设定与特定的质量和离子的电荷有关，能应用于鉴定。

MS 最普遍的应用如下，先用胰蛋白酶消化蛋白质，产生一系列的多肽，通过与从基因组序列预测到的所有蛋白质都可以产生的胰蛋白酶水解的多肽比较，MS 可以精确地测定每种多肽的相对分子质量。这将能鉴定一些蛋白质，包括二维胶上的斑点，甚至 SDS-PAGE 胶中的条带。

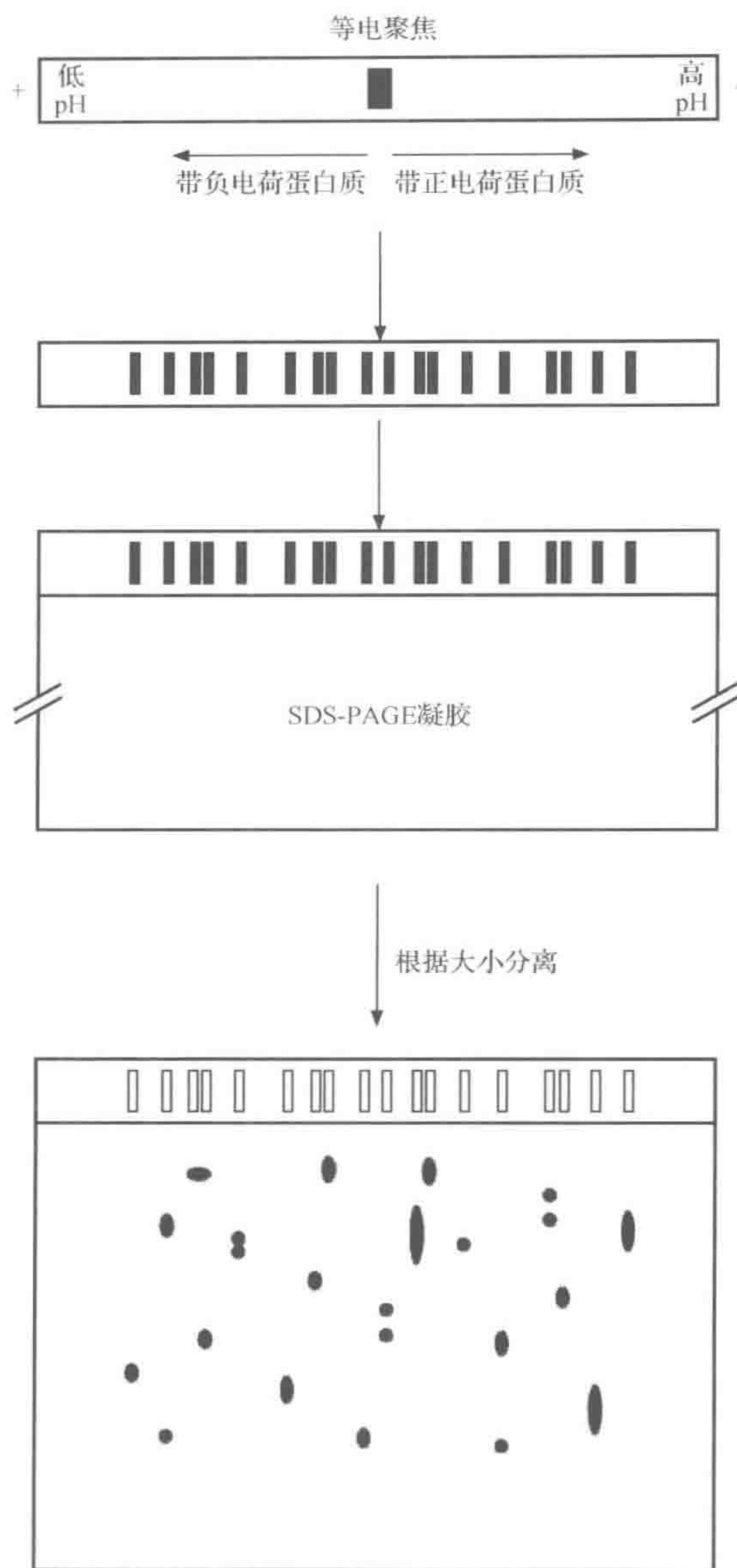


图 10.14 二维聚丙烯酰胺凝胶电泳

10.6 代谢组学

蛋白质组学分析方法允许在整个机体水平进行，同时测定蛋白质的丰度和分布。当这些蛋白质是酶类时，如果一种特殊酶的丰度改变，则可假定酶底物的水平及其产物的相应改变，就可以给出催化反应产物的信息。然而，蛋白质组学仅仅给出来自部分代谢及分解代谢途径的小分子水平的间接信息，更直接的方法是测定这些小分子的水平，代谢水平的全面性研究被称为代谢组学。由于高通量代谢谱的研究正在进行中，这是一种相对新

的学科和技术。依据其化学和物理特性,可以用气相色谱、高压液相色谱和毛细管电泳等方法分离代谢物,然后通过质谱分析的方法鉴定这些分子。

10.7 系统生物学和合成基因组学

10.7.1 系统生物学

上述的全部技术手段提供了大量的生物学信息,涉及细菌细胞内的不同水平,从DNA到mRNA、蛋白质甚至到代谢物。系统生物学就是从"组学"计划寻求统一信息而自然形成的一个新领域,以至于某一天能了解生物系统以整体发挥功能。其总体目标为了解DNA、mRNA、蛋白质、感觉网络和生物化学途径等不同层次间的相互作用及环境对细菌的影响。这包括所有细胞内已知酶、调节蛋白的水平和活性数据的数学模型的研究,因此可预测、检测抑制和改变特定的酶或删除某种基因的影响。未来几年,这可能再次使得研究细菌的方式产生巨大的变化。

10.7.2 合成基因组学

一旦经过注释,基因组序列可提供细菌不同组成部分的一部分列表。这些序列可给出一个特定DNA序列(与功能相关的一个基因或基因簇)的所有信息,可以严重影响细菌特性。如果将细菌基因组序列设想为改造细菌的指导手册,那么去想象产生完全合成的基因组,并产生合乎要求的具有所需特征的人造细菌并非难事。

科学家已开始着手人工合成细菌基因组的第一步,有望于制造的"设计细菌"可能生产生物燃料、净化污染点或人工生产新抗生素。制造合成基因组的关键步骤之一是发现对非依赖性生命所必需的最小基因组,这将形成基本基因组,所需的功能可以被添加到基本基因组。迄今为止,在不需要活宿主的培养条件下能够生长的最小基因组是生殖道支原体的基因组。因此,它可能包含了仅能维持非活体生长的数量最少的基因。因此,2005年,该细菌学成为了试图发现在不杀死生物体的前提下可以敲除多少基因的研究的开始。生殖器支原体基因组共有482个基因,其中382个是必需的(这些基因失活将杀死细菌),因此这就可以确定为维持自由生长细菌所必需的最小基因组。这代表了合成基因组计划所迈出的重要的第一步。

如果合成基因组学要发挥作用,合成基因组必须被移植到宿主细胞,然后能够重编程该细胞,使其能够具备所替代的基因组编码的特性。2007年,一个研究小组成功地将丝状支原体(*Mycoplasma mycoides*)的全基因组移植进入一种细菌山羊支原体(*Mycoplasma capricolum*)的相似的细菌细胞内,从而完成了产生合成基因组的下一个关键步骤。

为了制备一个合成基因组,制造将要组成基因组的DNA是必需的过程。2008年,通过合成和组装生殖器支原体细菌的582 970bp,科学家构建了最大的人造DNA结构,从而在这方面获得了突破性的进展。在实验室条件下,化学合成101个片段,每个片段5000～7000bp,覆盖了整个基因组。然后将这些片段导入酿酒酵母中进行组装,应用同源重组技术以较小的片段构建了整个细菌染色体。这一关键的基础步骤已完成,产生世

界首例合成生物仅仅是个时间问题。

10.8 结论

本书的最后一章不仅是对现有技术领域的简短综述，在某种程度上，分子遗传学增强了对于细菌基础生物学的理解（不必说已发生在高等生物中的革命）。在最后一章中提到的高通量技术或者全局技术，正在重新定义所实践的细菌遗传学和微生物学的技术方法。值得注意的技术进展意味着1天内可能获得细菌基因序列（高通量DNA测序），在8h内研究高达10 000个基因的表达变化（微阵列），通过6h的一次性筛选，2D-PAGE系统可以监测到超过3000个蛋白质的表达。因此，今天的科学家不再花费数年的时间来单独定义分离子中的单一蛋白和基因，却可以同时定义几百个不同的生物分子。这些进展不仅从根本性上冲击这些生物的分子生物学，而且正在改变着看待和研究它们的生理学、生物化学、病原学及进化的方式，没有任何例外。然而，为了解读这些信息，还要回溯到细胞作为一个整体的行为及其与环境的相互作用。这是多个不同领域的合成，诞生了系统生物学的研究目标，这些生物学数据的整合将是未来研究所面对的主要挑战。

A 补充书目

注释

列表只是选择性地列出了现有书目中最有用的一些。在一个像分子遗传学这样快速发展的领域，信息更新迅速。在某些情况下，语言的细微变化或概念的改变，或一个实验性的解释会为人们所接受，虽然它并不总是错的。虽然本书已经使用了最新版本，但作者仍在设法找到更新版本的教科书。本书涵盖了一些依然有价值的旧版本的书，而且尽可能参考了之前发表的文章和原作来补充这个列表。由于这些新的著作比教科书更新更快，本书没有收录这些内容。

通用书目和基础分子生物学

B. Lewin,《基因Ⅸ》(2007),Jones and Bartlett 出版社出版。该书可以作为进一步参考的首选，但是读者会发现比本书更有深度。这本书包含的范围很广，但对于细菌和真核生物的研究都非常有用。

C. R. Calladine 和 H. R. Drew,《认识 DNA：分子及其作用原理》(第三版)(2004),Academic 出版社。该书的特色在于，从看似简单的方式开始，进而阐释与 DNA 结构和功能相关的诸多复杂概念，尤其是超螺旋结构。强烈推荐这本书作为一些较浅地描述 DNA 结构的分子生物学著作的补充。

B. Alberts 等,《细胞分子生物学》(第五版)(2007),Garland Science 出版社。学术价值很高，但不方便随身携带。该书虽然主要与真核生物有关，但对于某些分子方面也极具参考价值。

H. Lodish 等,《细胞分子生物学》(2007),W. H. Freeman 出版社。这是另一本主要针对真核细胞生物学的书，但该书还包括了许多的相关性信息，以及只与细菌有关的内容，在更广泛的分子领域会更具参考价值。

更多关于细菌的著作

L. Snyder 和 W. Champness,《细菌分子遗传学》(第三版)(2007),美国微生物学会。细菌遗传学进一步参考著作的首选。

S. Baumberg(ed.),《原核基因表达》(1999),牛津大学出版社。该书由多位作家参与

编写，涉及有关特定选题的详细资料，但并非相关内容的简单整合。该书是本书内容的提升。

M. T. Madigan，J. M. Martinko，P. V. Dunlap 和 D. P. Clark，《微生物生物学》（第十二版）（2008），Benjamin Cummings 出版社。可以从该书中查阅一些生物学背景知识。

基因克隆

J. W. Dale 和 M. von Schantz，《从基因到基因组》（第二版）（2007），Wiley-Blackwell 出版社。通过该书，可以更深入地了解 DNA 重组技术，并进入真核生物领域。

T. A. Brown，《基因克隆与 DNA 分析：介绍》（第五版）（2006），Wiley-Blackwell 出版社。这是一本不错的入门著作。

S. B. Primrose 和 R. M. Twyman，《基因操作和基因组学原理》（第七版）（2006），Wiley-Blackwell 出版社。推荐供更高层次的学生使用。

B. R. Glick，J. J. Pasternak 和 C. L. Patten，《分子生物技术：重组 DNA 分子的原理及应用》（第四版）（2009），美国微生物学会。该书更关注于分子技术的广泛应用。

专题

R. W. Hendrix 等，《Lambda Ⅱ》（1983），冷泉港实验室。尽管这本书出版时间较早，但它仍然是一本非常珍贵的关于 λ 噬菌体的参考书。本书仅针对这个领域更加专业的学生。

A. A. Salyers 和 D. D. Whitt，《细菌致病机理》（第二版）（2002），美国微生物学会。在细菌的致病机制方面极具参考价值。

A. M. Campbell 和 L. J. Heyer，《基因组学，蛋白质组学，生物信息学的发现》（第二版）（2007），冷泉港实验室。该书将基因组学的应用延伸到细菌以外的领域，对掌握许多以前应用过的方法非常有用。

A. L. esk，《生物信息学简介》（第二版）（2008），牛津大学出版社。面向研究生的生物信息学的介绍。

关注

W. Hayes，《细菌及其病毒的遗传学》（第二版）（2008），Blackwell Scientific 出版社。作为一本经典书目，对于那些对本学科发展历史感兴趣的读者，该书值得一读。

网站

这里只列出了一小部分挑选出的，特别是那些与基因序列有关的有用的网址。尽管这些网站的主页或许是相当稳定的，但是要注意这些网站地址可能会发生改变。

网址	机构	包括的工具
www. ebi. ac. uk	欧洲生物信息研究所	EMBL 等数据库 BLAST 等工具
www. ncbi. nlm. nih. gov	美国国家生物信息技术中心(NCBI)	GenBank 数据库及搜索和比较的工具
www. sanger. ac. uk	桑格研究院(WTSI)	基因组序列数据和工具,尤其是关于病原体的信息
www. jcvi. org	克雷格·文特尔研究所	基因组序列计划,包括综合微生物资源,包括所有的细菌基因序列
www. genomesonline. org		基因组序列的列表

B 常用缩写

氨基酸符号

	中文名称	三字符	一字符
Alanine	丙氨酸	Ala	A
Arginine	精氨酸	Arg	R
Asparagine	天冬酰胺	Asn	N
Aspartate	天冬氨酸	Asp	D
Cysteine	半胱氨酸	Cys	C
Glutamate	谷氨酸	Glu	E
Glutamine	谷氨酰胺	Gln	Q
Glycine	甘氨酸	Gly	G
Histidine	组氨酸	His	H
Isoleucine	异亮氨酸	Ile	I
Leucine	亮氨酸	Leu	L
Lysine	赖氨酸	Lys	K
Methionine	甲硫氨酸	Met	M
Phenylalanine	苯丙氨酸	Phe	F
Proline	脯氨酸	Pro	P
Serine	丝氨酸	Ser	S
Threonine	苏氨酸	Thr	T
Tryptophan	色氨酸	Trp	W
Tyrosine	酪氨酸	Tyr	Y
Valine	缬氨酸	Val	V

其他缩写

A	腺嘌呤
AFLP	扩增片段长度多态性
BAC	细菌人工染色体
BCG	卡介苗
BLAST	局部序列排比对检索基本工具
5BU	5-溴尿嘧啶
C	胞嘧啶
cAMP	环腺苷酸
CAP	环腺苷酸结合蛋白质
CAS	CRISPR 关联序列
ccc	共价闭环(质粒 DNA)
cDNA	互补 DNA
CHEF	等强度均一电场电泳
ChIP	染色质免疫沉淀
CRISPR	有规律间隔的短回文重复序列
CRP	环腺苷酸受体蛋白(CAP 的同义词)
ddNTP	双脱氧核苷三磷酸,包括 ddATP、ddCTP、ddGTP、ddTTP
DGGE	变性梯度凝胶电泳
dNTP	脱氧核苷三磷酸,dATP、dCTP、dGTP、dTTP 其中一种或混合物
DR	同向重复
ds	双链(DNA 或 RNA)
ECF	胞质外的 σ 功能因子
EMS	甲基磺酸乙酯
e. o. p	感染效率(噬菌体)
FIGE	反转电场凝胶电泳
fMet	甲酰甲硫氨酸
G	鸟嘌呤
GFP	绿色荧光蛋白
GSP	一般分泌途径
Hfr	高频重组
HPK	组氨酸蛋白激酶
IEF	等电聚焦
IPTG	异丙基-β-D-硫代半乳糖苷
IR	反向重复
IS	插入序列
IVET	体内表达技术
kb	千碱基
kDa	千道尔顿

Mb	兆碱基
MLST	多位点序列分型
MNNG	1-甲基-3-硝基-1-亚硝基胍
mRNA	信使 RNA
MS	质谱法
NTP	ATP、GTP、CTP、UTP 其中的一种或者其混合物
oc	开环 DNA(带切口的质粒 DNA)
ORF	可读框
PAGE	聚丙烯酰胺凝胶电泳
PCR	聚合酶链反应
PEG	聚乙二醇
PFGE	脉冲电场凝胶电泳
pfu	噬斑形成单位
pI	等电点(蛋白质没有净电荷时的 pH)
PP_i	焦磷酸盐
RBS	核糖体结合位点
RF	复制型
RFLP	限制性片段长度多态性
RR	反应调控子
rRNA	核糖体 RNA
RT-PCR	反转录聚合酶链反应
SDS	十二烷基硫酸钠
ss	单链的 DNA 或 RNA
SSB	单链 DNA 结合蛋白
STM	信号标签诱变技术
T	胸腺嘧啶
TGGE	温度梯度凝胶电泳
Ti	肿瘤诱发
Tn	转座子
TRAP	色氨酸 RNA 结合衰减子蛋白
TraSH	转座子位点杂交
tRNA	转移 RNA
U	尿嘧啶
UV	紫外辐射
VNTR	数目可变串联重复序列
X-gal	5-溴-4-氯-3-吲哚-β-D-半乳糖苷

C 词 汇 表

本书涉及的关于酶和基因的信息见附录 D 和附录 E。

Adaptive mutation	适应性突变
Allele	等位基因
Allelic replacement	等位基因置换
Allosteric effect	变构效应
Amber mutation	琥珀突变
Aminoglycosides	氨基糖苷类
Ampicillin	氨苄青霉素
Amplicon	扩增子
Amplification	放大
Analogues	类似物
Annealing	退火
Annotation	注释
Anticodon	反密码子
Antimetabolite	抗代谢物
Antisense RNA	反义 RNA
Anti-σ-factor	抗 σ 因子
Anti-terminator	抗终止子
Aptamer	适配体
Attenuation	减毒作用
Autogenous control	自控
Autoradiography	放射自显影
Auxotroph	营养缺陷型
Back mutation	回复突变
Bacterial artificial chromosome	细菌人工染色体
Bacteriocin	细菌素

Bacteriophage	噬菌体
β-Lactams	β-内酰胺类抗生素
Bioinformatics	生物信息学
Capsid	衣壳
Cassette	盒
Catabolite repression	分解代谢物阻遏
cDNA library	cDNA 文库
Chaperone	分子伴侣
Chemotaxis	趋化性
Chloramphenicol	氯霉素
Chromatin	核染色质
Chromosome walking	染色体步移
Circular permutation	循环交替变换
Cis-acting	顺式作用
Cistron	顺反子
Clonal	无性(繁殖)系的
Clone	克隆
Cloning	无性繁殖
Codon	密码子
Codon usage	密码子选择
Cohesive ends	黏性末端(黏端)
Cointegrate	共合体
Colicin	大肠(杆)菌素
Competent	感受态
Complementary DNA	互补 DNA
Complementary strand	互补链
Complementation	互补作用
Composite transposon	复合转座子
Conditional mutant	条件突变体
Conjugation	共轭,接合
Conjugative transposon	接合转座子
Consensus	共有序列

Constitutive	组成型
Contig	重叠群
Copy number	拷贝数
cos site	黏性位点
Cosmid	黏粒
Co-transduction	共转导
Co-transformation	共转化
CRISPR(clustered regularly interspersed short palindromic repeats)	成簇的有规律的间隔短回文序列
Cross-talking	信号对答
Curing	固化
Denaturation	变性
DGGE(denaturing gradient gel electrophoresis)	变性梯度凝胶电泳
Diauxic growth	双峰生长
Direct repeat	同向重复(序列)
Directed mutation	诱发突变
Distal	末端的
Domain	结构域
Dyad symmetry	二重对称
Early genes	早期基因
Efficiency of plating	感染效率
Electrophoresis	电泳
Electroporation	电穿孔
Epigenetic	表观遗传
Episome	游离基因
Error-prone repair	易错修复
Eukaryote	真核生物
Excision repair	切除修复
Exon	外显子
Expression vector	表达型载体
Extracytoplasmic function σ factors	细胞质外的功能性 σ 因子
Plasmid	质粒

Feedback	反馈
Fimbria	菌毛
Flagellum	鞭毛
Fluctuation test	彷徨变异实验
Frameshift mutation	移码突变
Fusion protein	融合蛋白
Gene conversion	基因转换
Gene knockout	基因敲除
Gene library	基因文库
Gene replacement	基因置换
Genetic map	遗传图
Genome	基因组
Genome sequencing	基因组测序
Genotype	基因型
Global regulation	全局调节
Gratuitous inducer	义务诱导物
Hairpin	发夹
Headfull	头容纳量
Hemimethylated	半甲基化的
Heteroduplex	异源双链
Hfr(high frequency of recombination)	高频重组
Holliday junction	Holliday 连接体
Homologous	同源的
Homologous recombination	同源重组
Homologues	同系物
Homopolymer	均聚物
Horizontal gene transfer	水平基因转移
Host range	宿主范围
Hybridization	杂交
Hydrophilic	亲水的
Hydrophobic	疏水的
Illegitimate recombination	异常重组

In vitro packaging	体外包装
Incompatibility	不相容性
Indel	缺失多态性
Induction	诱导
Insertion sequence	插入序列
Insertion vector	插入载体
Insertional inactivation	插入失活
Integron	整合子
Intercalation	插入
Intermolecular	分子间的
Intramolecular	分子内的
Intron	内含子
Inverse PCR	反向 PCR
Inversion	倒位
Inverted repeat	反向重复序列
ISCR element	ISCR 元件
Isoelectric point	等电点
Isogenic	同基因的
Iteron	重复子
IVET(*in vivo* expression technology)	体内表达技术
Kanamycin	卡那霉素
Kilobase(kb)	千碱基
Knockout	敲除
Lagging strand	后随链
Leaky	漏的,有漏洞的
Ligand	配体
Ligation	连接
Linking number	连接数
Lysogeny	溶源性
Lytic cycle	裂解周期
Mapping	图谱
Mass spectrometer	质谱仪

Massively parallel sequencing	大规模平行测序
Melting	融化
Messenger RNA(mRNA)	信使 RNA
Metabolomics	代谢组学
Microarray	微阵列
Minicells	微细胞
Minus strand	负链
Mismatch repair	错配修复
Mobilization	动员
Modification	修正
Mosaic genes	嵌合基因
Motif	基序
Multilocus sequence typing(MLST)	多位点序列分型
Multiple cloning site	多克隆位点
Mutagenesis	突变形成,变异发生
Mutant	突变体
Mutation	突变
Nalidixic acid	萘啶酸,萘啶酮酸
Nick	缺口
Nonsense mutation	无义突变
Northern blot	RNA 印迹
Okazaki fragment	冈崎片段
Oligonucleotide	寡核苷酸
Open reading frame(ORF)	可读框
Operator	操纵基因
Operon	操纵子
Ordered library	有序文库
Origin of replication	复制起始点
Orthologues	直系同源
Packaging	包装
Paralogues	旁系同源物
Partial diploid	部分二倍体

Partitioning	分区
Pathogenicity island	毒力岛,致病岛
Penicillin enrichment	富集青霉素
Penicillin	青霉素
Permissive host/conditions	条件性寄主
Phage	噬菌体
Phage conversion	噬菌体转换
Phage display	噬菌体展示
Phage therapy	噬菌体疗法
Phage typing	噬菌体分型
Phage variation	噬菌体变异
Phenotype	表型
Pheromone	信息素
Phosphorelay	磷酸化
Photoreactivation	光复活作用
Phylogenetic tree	系统树
Physical map	物理图
Pilus(pl. pili)	菌毛
Plaque	噬菌斑
Plasmid	质粒
Plasmid addiction	质粒依赖型
Plasticity	可塑性
Plus strand	正链
Point mutation	点突变
Polarity	极性
Polycistronic mRNA	多顺反子 mRNA
Polymerase chain reaction	聚合酶链反应
Polymorphism	多态性
Post-replication repair	复制后修复
Post-translational modification	翻译后修饰
Pribnow box	普里布诺框
Primary structure	一级结构

Primer	引物
Probe	探针
Prokaryote	原核生物
Promoter	启动子
Proofreading	校正
Prophage	原噬菌体
Protein engineering	蛋白质工程
Proteome	蛋白质组
Protoplast	原生质(体)
Prototroph	原养型,原养生物
Proximal	近端的
Pseudogene	假基因,拟基因,伪基因
Pulsed-field gel electrophoresis	脉冲电场凝胶电泳
Purine	嘌呤
Pyrimidine	嘧啶碱基
Pyrimidine dimer	嘧啶二聚体
Pyrosequencing	焦磷酸测序
Quantitative PCR	定量 PCR
Recombinant	重组子
Recombination repair	重组修复
Recombination	基因重组
Regulatory RNA	调控 RNA
Regulon	调节子
Relaxation	松弛,缓和
Relaxed plasmid	松弛质粒
Replacement vector	置换型载体
Replica plating	复制平皿(板)培养法
Replication	复制
Replicative form	复制型
Replicon	复制子
Reporter gene	报道基因
Repression	抑制

Resolution	分辨率
Restriction-fragment length polymorphism(RFLP)	限制性片段长度多态性
Restriction mapping	限制作图
Restriction	限制
Reverse genetics	反向遗传学
Reverse transcription	反转录
Reversion	逆转
Ribosomal frame shifting	核糖体移码
Ribosome-binding site(RBS)	核糖体结合位点
Riboswitch	核糖开关
Rifampicin	利福平
RNA interference	RNA 干扰
RNA thermometer	RNA 温度计
Rolling-circle replication	滚环复制
RT-PCR(reverse-transcript PCR)	反转录聚合酶链反应
Sanger sequencing	桑格法测序
Scaffolding	支架
SDS-PAGE	十二烷基硫酸钠聚丙烯酰胺凝胶电泳
Secondary structure	二级结构
Segregation	分离
Selectable marker	选择标记
Sequence polymorphism	序列多态性
Shine-Dalgarno sequence	SD 序列
Shotgun cloning	鸟枪法克隆
Shotgun sequencing	鸟枪法测序
Shuttle vector	穿梭载体
Factor(sigma factor)	σ 因子
Signal peptide	信号肽
Signal transduction	信号转导
Signature-tagged mutagenesis	信号标签诱变技术
Silent mutation	沉默突变
Site-directed mutagenesis	定点诱变

Site-special recombination	位点特异性重组
Slipped-strand mispairing	滑动链错配
SOS response	SOS 反应
Southern blot	DNA 印迹
Specialized transduction	局限性转导
Splicing	剪接
Start codon	起始密码子
Stem-loop structure	茎环结构
Stop codon	终止密码子
Streptomycin	链霉素
Stringency	严谨性
Stringent response	严谨性反应
Structural genes	结构基因
Stuffer fragment	填充片段
Supercoiling	超螺旋
Superinfection immunity	超感染免疫性
Suppression	抑制
Synonymous codons	同义密码子
Synteny	同线性
Systems biology	系统生物学
T box	T 盒
Tag	标签
Tandem repeat	串联重复
Tautomerism	互变异构
Taxonomy	分类学
Temperate	温和的
Template	模板
Terminal redundancy	末端冗余
Termination	终止
Terminator	终止子
Tertiary structure	三级结构
Tetracycline	四环素

TGGE(temperature-gradient gel electrophoresis)	温度梯度凝胶电泳
Theta replication	θ型复制
Thymine dimmer	胸腺嘧啶
Ti plasmids	Ti 质粒
Titre	滴度
Trans-acting	反式作用
Transconjugant	转接合子
Transcription	转录
Transcriptome	转录组
Transduction	转导
Transfection	转染
Transformation	转化
Translation	翻译
Transposon	转座子
Transposon mutagenesis	转座子诱变
Transponson-site hybridization	转座子插入位点杂交
Twist	弯曲
Two-component regulation	双组分调节
Two-dimensional gel electrophoresis	双向电泳技术
UP element	UP 元素
Vector	载体
Virulent	毒力的
Variable-number tandem repeat(VNTR)	可变数目串联重复
Western blot	蛋白质印迹
Wobble	摆动
Writhe	使扭曲缠绕
X-gal	5-溴-4-氯-3-吲哚-β-D 半乳糖苷
Z DNA	Z 型 DNA
Zygotic induction	接合子诱导

D 酶及其他蛋白质

这个列表包含了书中提到的一些酶及其他蛋白质的简要介绍。

腺苷酸环化酶：产生 cAMP。

碱性磷酸酶：(a)在基因克隆中用来去掉 DNA 分子 C 端的磷酸基团；(b)作为识别分泌信号的报道基因。

氨酰 tRNA 合成酶：其功能是用合适的氨基酸补充特定的 tRNA。

阿糖胞苷：阿拉伯糖操纵子的调控物。

天冬氨酸激酶：催化氨基酸生物合成的一些初始步骤，是一个关键的调控步骤。

β-半乳糖苷酶：将乳糖分解成为葡萄糖和半乳糖。乳糖发酵的第一步。

β-内酰胺酶：水解青霉素和头孢菌素核内的 β-内酰胺键，与青霉素抗性相关。

环磷酸腺苷受体蛋白(CRP)或代谢活化蛋白(CAP)：与环磷酸腺苷相关，活化乳糖操纵子的转录。它是一个整体调节子，负责某些形式的分解代谢产物阻遏。

氯霉素转乙酰酶(CAT)：通过抑制抗生素的活性产生氯霉素抗性。

cⅠ抑制因子：抑制溶源态中的 λ 噬菌体的蛋白质因子。

Cro：λ 噬菌体蛋白在促细胞溶解的和溶源的转换中有关键性的作用，为抵抗 cⅠ因子所必需。

CsrA：对营养缺乏做出反应的调控蛋白。与 mRNA 结合，影响其稳定性和翻译。

Dam：腺苷脱氧甲基化酶。

DnaA：染色体复制起始需要的起始因子。

DNAligase：DNA 连接酶。

DNA 聚合酶Ⅰ：主要作为一个修复聚合酶，作用于单链缺口；也参与修复 DNA 复制过程中形成后随链时的缺口。同时具有包括 C 端到 N 端和 N 端到 C 端外切核酸酶活性。

DNA 聚合酶Ⅲ：主要的 DNA 复制聚合酶。

内切核酸酶：可以在内部切断 DNA 的一种酶。

外切核酸酶：从 DNA 片段末端切除核苷酸的一种酶。从 C 端到 N 端的外切核酸酶从 C 端切除核苷酸，而从 N 端到 C 端的外切核酸酶从 N 端切除核苷酸。

反向刺激因子(FIS)：一个整体调节子，通过折叠 DNA 改变一系列基因的表达。

FlgM：反-σ 因子，调控沙门菌属的鞭毛的形成。

FNR：一个整体调节子，控制许多微需氧基因的扩增。

谷氨酸脱氢酶：催化 α-酮戊二酸和氨反应，生成谷氨酸，在氨的同化作用中起重要的作用。

绿色荧光蛋白:本质上是荧光报道蛋白。

螺旋酶:一种特殊的拓扑异构酶,向 DNA 引入负超螺旋。

解旋酶:解开 DNA 链,如在接合质粒转移中发挥作用。

Hin:重组酶的特异性位点,作用于转变沙门氏菌属的相变异的控制因素。

抗组蛋白样拟核结构蛋白(H-NS):一个 DNA 结合蛋白,影响 NDA 的局部拓扑结构,导致基因的表达。

组氨酸蛋白激酶(HPK):当受到刺激时,作为回应,组氨酸残基自身磷酸化,形成的磷酸盐被转运到一个反应调节子。它们共同组成一个双组分的调节系统。

HU:DNA 结合蛋白,影响 DNA 局部的拓扑结构,导致基因表达。

整合酶:整合酶通过位点特异性重组将 λDNA 整合入染色体。整合酶的相关蛋白质催化其他成分的整合,如接合型转座子。参见 Xis。

整合作用宿主因子(IHF):整体调控子,通过弯曲 DNA 来改变一系列基因的表达。也被用于将 λDNA 有效整合入染色体。

激酶:磷酸化酶。

LacI:乳糖操纵子的阻遏物。

LexA:SOS 反应的调节子。

连接酶:密封双链 DNA 中的单链缺口,也参与基因克隆中重组 DNA 分子的结构形成。

萤光素酶:催化发光反应,作为一个报道信号。

LuxI:高丝氨酸内酯合成酶,负责一个涉及群体效应的信使的合成。

LuxR:转录调节子,对高丝氨酸内酯的水平做出反应。

OxyR:整体调节蛋白,对氧化应激做出反应。

光裂合酶:负责紫外线诱导的嘧啶二聚体的光活化移除。

菌毛蛋白:菌毛结构亚基。

多核苷酸激酶:能将 ATP 的磷酸基团转移到 DNA 或 RNA 的 C 端。

引发酶:一种特殊的 RNA 聚合酶,在 DNA 合成时需要该酶产生一段短的引物。

RecA:同源重组以及 DNA 修复中的关键蛋白质。

RecBCD:涉及同源重组的多功能酶。

复制酶:RNA 引导的 RNA 聚合酶,用于某些 RNA 病毒的复制。

反应调控子(RR):两个元件组成的监管体系中的一部分,该体系结合在 DNA 上,使该段基因的转录启动或关闭。被组氨酸蛋白激酶的磷酸化激活。

解离酶:催化位点特异性重组来解离转位过程中的中间共合体使其成为两个独立的分子。

限制酶:识别并剪切 DNA 的特殊位点。最广泛使用的是限制酶Ⅱ,在识别位点内切割。

反转录酶:RNA 引导的 DNA 多聚合酶,使用 mRNA 模板合成 DNA。

RNA 酶:降解 RNA 分子。

RNA 聚合酶:通过 DNA 模板合成 RNA。

核糖核酸酶:水解 RNA-DNA 杂合体的一种特异性 RNA 酶,涉及 ColE Ⅰ 样质粒的复制。

Ruv:涉及同源重组中 Holliday 交叉反应的某些蛋白质。

σ 因子:RNA 聚合酶的可分离部分,负责启动子特异性。

单链 DNA 结合蛋白:稳定单链 DNA,如在 DNA 复制过程中。

拓扑异构酶:能改变 DNA 结构的一类酶,如能通过改变 DNA 螺旋或超螺旋的度改变 DNA 结构。

转座酶:催化转座的起始步骤。

色氨酸-RNA 结合弱化子蛋白:负责杆菌色氨酸操纵子的弱化。

XerC,D:在 *cer* 位点通过重组分解质粒二聚体。

Xis:与整合酶互相作用促进 λ 前噬菌从体染色体中切离。相关蛋白与其他成分如接合型转座子有类似的功能。

E 基　　因

这是书中所涉及的简化了的基因/操纵子的摘录。值得注意的是，在许多情况下，完成指定的功能需要多种基因的参与，而由于特异性 tRNA 的产生，某些具有相同 3 个字母的基因的功能却不同。

基因

基因	基因产物/表型特征
aac	氨基糖苷转乙酰酶(庆大霉素抗性)
aad	氨基糖苷类果糖(链霉素抗性)
amp	内酰胺酶(氨苄青霉素抗性)
aph	氨基糖苷磷酸转移酶(卡那霉素抗性)
ara	利用阿拉伯糖
araC	阿拉伯糖操纵子的调控子
arg	精氨酸生物合成
bio	生物素合成
bla	氨苄青霉素抗性(β-内酰胺酶)
bld	在天蓝色链霉菌，突变体不能产生气生菌丝
bom	ColE1 的移动的缺口位点
cat	氯霉素转乙酰酶(氯霉素抗性)
ccd	游离质粒分离后死亡
cer	多长度质粒串联体降解的重组位点，导致不稳定性的缺失
cml	同 *cat*
col	产生大肠杆菌素
dhfr	二氢叶酸还原酶
dsrA	调节性 RNA，调节 *hns* 和 *rpoS*
fimA	菌毛的结构亚基
ftsZ	启动细胞分裂，条件突变株在限制性温度下产生细丝

续表

基因	基因产物/表型特征
gal	半乳糖发酵
hin	H2 操纵子(沙门氏菌鞭毛抗原)的调节元件的倒位
his	组氨酸合成
hok/sok	游离质粒分离后死亡
ilv	合成异亮氨酸和缬氨酸
inc	质粒不兼容性
kan	卡那霉素抗性
lac	半乳糖发酵,由β-半乳糖苷酶、透(性)酶和转乙酰酶组成的操纵子
lac Ⅰ	半乳糖操纵子的阻遏物,组成型突变
leu	亮氨酸的合成
lexA	lexA 蛋白阻遏 SOS 应答的基因
lys	赖氨酸合成
mer	汞离子抗性
micF	反义 RNA,调节 *ompF*(外膜孔道蛋白)
min	在细胞分裂时调节隔膜形成的位置,突变体产生缺少 DNA 的微细胞
mob	质粒转移所需的核酸酶
mucAB	质粒传播的基因,能增加紫外线辐射的诱变效应
nal	萘啶酸抗性
nif	固氮作用
nifA	*nif* 基因的正调节子
ompF	外膜蛋白(一类细胞外膜孔道蛋白)
oriT	一个质粒转移的起始位点
oriV	复制的起始位点
par	细胞分裂时质粒的分配
pheA	苯丙氨酸合成
pil	菌毛蛋白,如淋球菌菌毛的亚单位
pro	脯氨酸合成
recA	recA 蛋白介导非特异性重组、修复过程和噬菌体诱导

续表

基因	基因产物/表型特征
recB,*recC*	编码具有外切酶和内切酶活性的多功能酶,用于 recA 介导的重组
rep	质粒复制
res	转座中的分解位点
rpoB	RNA 聚合酶的 β 亚基;某些突变体能产生利福平抗性
rpoS	固定相 σ 因子
rpsL	核糖体蛋白 S10,链霉素抗性(染色体突变)
sig	σ 因子,因为 σ 因子是 RNA 聚合酶的一个(可解离的)亚基,通常表示为 *rpo*,如 *rpoS*
spo ⅡA	控制枯草芽孢杆菌孢子形成的一个时期,*spo* Ⅱ*AC* 编码 σ 因子,而 *spo* Ⅱ*AB* 和 spoⅡAA 分别编码一种反 σ 因子和 spoⅡAB 的拮抗子
str	链霉素抗性(作为一个染色体标记,通常为 *rpsL*;位于质粒或转座子上,通常是氨基糖苷类修饰酶,如 AadA)
sul	硫胺类药剂抗性
ter	复制终止位点
tet	四环素抗性
thr	苏氨酸合成
thyA	胸苷酸合成酶,胸腺嘧啶/胸腺嘧啶脱氧核苷生成
tnpA	转座子酶
tnpR	解离酶
tra	转位,用于接合质粒转移的一组基因
trp	色氨酸生物合成
trpA,*trpB*	色氨酸合成酶亚基
trpR	色氨酸操纵子的阻遏物
umuC,*umuD*	紫外线突变,这些基因的缺陷会增强对紫外线辐射造成的致命效果的敏感性,但突变的发生率会降低
uvr	紫外线修复,几个基因(*uvrA*,*uvrB*,*uvrC*)编码负责胸腺嘧啶二聚体去除的核酸酶
whiG	位于天蓝色链霉菌,编码孢子形成所需的 σ 因子

λ基因的选择

基因	基因产物/表型特征
att	将 λDNA 整合入染色体的位点；噬菌体序列被叫做 *attP*，其在染色体中被称为 *attB* 或 *att*λ；整合完成后，原噬菌体这一位点的左端与右端分别被称为 *attL* 或 *attR*
*c*Ⅰ	CI 阻遏物，对于溶源是必需的，结合到 O_L 和 O_R
*c*Ⅱ，*c*Ⅲ	由 P_E 激活，阻遏物最初合成所需的辅助基因
cos	黏性末端位点，包装过程中不对称切割的位置
cro	Cro 蛋白本质上起抗阻遏物的作用，对抗 CⅠ阻遏物的作用
int	将 λDNA 整合入染色体，与 Xis 一起，删除原噬菌体
N	抗终止子；允许通读转录终止位点，造成延迟早期基因的表达
O_L，O_R	CI 和 Cro 结合的操纵位点，调控 P_L 和 P_R 的表达
P_E	溶源形成过程中 CⅠ表达的启动子，由 CⅡ、CⅢ激活
P_L，P_R	分别是主要的早期转录物向左和向右的启动子
P_M	产生溶源后，*c*Ⅰ表达的维护子
$P_{R'}$	主要的晚期转录物的启动子
Q	被需要于来自 $P_{R'}$ 的晚期基因的表达
S	裂解
t_{R1}，t_{R2}	来自 P_R 即时早期转录物的终止位点，N 蛋白的允许通读的抗终止活性
t_{R3}	来自 P_R 延迟的早期转录物的终止位点
xis	与 Int 一起，当诱导裂解周期时原噬菌体的切除

F 标准遗传密码

UUU Phe 苯丙氨酸	UCU Ser 丝氨酸	UCC Ser 丝氨酸	UGU Cys 半胱氨酸
UUC Phe 苯丙氨酸	UCC Ser 丝氨酸	UAC Tyr 酪氨酸	UGC Cys 半胱氨酸
UUA Leu 亮氨酸	UCA Ser 丝氨酸	UAA 终止	UGA 终止
UUG Leu 亮氨酸	UCG Ser 丝氨酸	UAG 终止	UGG Trp 色氨酸
CUU Leu 亮氨酸	CCU Pro 脯氨酸	CAU His 组氨酸	CGU Arg 精氨酸
CUC Leu 亮氨酸	CCC Pro 脯氨酸	CAC His 组氨酸	CGC Arg 精氨酸
CUA Leu 亮氨酸	CCA Pro 脯氨酸	CAA Gln 谷氨酰胺	CGA Arg 精氨酸
CUG Leu 亮氨酸	CCG Pro 脯氨酸	CAG Gln 谷氨酰胺	CGG Arg 精氨酸
AUU Ile 异亮氨酸	ACU Thr 苏氨酸	AAU Asn 天冬酰胺	AGU Ser 丝氨酸
AUC Ile 异亮氨酸	ACC Thr 苏氨酸	AAC Asn 天冬酰胺	AGC Ser 丝氨酸
AUA Ile 异亮氨酸	ACA Thr 苏氨酸	AAA Lys 赖氨酸	AGA Arg 精氨酸
AUG Met 甲硫氨酸	ACG Thr 苏氨酸	AAG Lys 赖氨酸	AGG Arg 精氨酸
GUU Val 缬氨酸	GCU Ala 丙氨酸	GAU Asp 天冬氨酸	GGU Gly 甘氨酸
GUC Val 缬氨酸	GCC Ala 丙氨酸	GAC Asp 天冬氨酸	GGC Gly 甘氨酸
GUA Val 缬氨酸	GCA Ala 丙氨酸	GAA Glu 谷氨酸	GGA Gly 甘氨酸
GUG Val 缬氨酸	GCG Ala 丙氨酸	GAG Glu 谷氨酸	GGG Gly 甘氨酸

列表中氨基酸的缩写已在附录 B 中

G 菌　种

请注意,除非另有指定,全文提及的菌为大肠杆菌。

细菌名称	缩写	其他名称	注释
根癌农杆菌	*A. tumefaciens*		植物病原菌,用于将 DNA 转入植物细胞
炭疽杆菌	*B. anthracis*		导致炭疽
枯草芽孢杆菌	*S. sublitis*		革兰氏阳性芽孢形成菌,广泛用于遗传学
类杆菌			革兰氏阴性肠道共生菌
百日咳杆菌	*B. pertussis*		导致百日咳
包柔氏螺旋体			与回归热和莱姆病有关的螺旋体
黄色短杆菌	*B. flavum*		现将其重新分类为谷氨酸棒状杆菌
布赫纳氏菌			蚜虫细菌共生生物的一个种属
胎儿弯曲杆菌	*C. fetus*		导致绵羊和牛的流产
空肠弯曲菌	*C. jejuni*		造成食源性传染疾病的一个主要原因
新月柄杆菌	*C. crescentus*		水生细菌;分化为茎细胞和游动细胞
肉毒梭菌	*C. botulinum*		导致肉毒杆菌中毒
破伤风杆菌	*C. tetani*		导致破伤风
白喉杆菌	*C. diphtheriae*		噬菌体编码的毒素导致白喉
谷氨酸棒状杆菌	*C. glutamicum*		用于工业生产,如生产氨基酸
粪肠球菌	*E. faecalis*	以前称为粪链球菌	革兰氏阳性肠道寄生菌,机会致病菌
胡萝卜软腐欧文氏菌	*E. carotovora*		植物病原菌
大肠杆菌	*E. coli*		遗传学研究的标准生物,革兰氏阴性肠道寄生菌,某些菌株导致严重疾病
流感嗜血杆菌	*H. influenzae*		呼吸道寄生菌,导致包括脑膜炎在内的严重疾病
螺杆菌	*H. pylori*		胃寄生菌,导致溃疡及其他疾病
百脉根中生根瘤菌	*M. loti*		因其具有固氮活性而对于农业十分重要

续表

细菌名称	缩写	其他名称	注释
微球菌种属			革兰氏阳性球菌,广泛存在于皮肤
麻风分枝杆菌	*M. leprae*	麻风菌	导致麻风病
副结核分枝杆菌	*M. paratuberculosis*		动物病原菌
结核分枝杆菌	*M. tuberculosis*	结核杆菌	导致人结核病
生殖支原体	*M. genitalium*		非常小的细菌,无细胞壁
黏液球菌	*M. xanthus*		土壤细菌,表现为协同行为
淋球菌	*N. gonorrhoeae*	淋菌	导致淋病
脑膜炎奈瑟菌	*N. meningiitidis*	脑膜炎球菌	导致脑膜炎和败血症
铜绿假单胞菌	*P. aeruginosa*	绿脓杆菌	革兰氏阴性机会致病菌,广泛用于遗传学研究
恶臭假单胞菌	*P. putida*		发现于水和土壤中的自由生活细菌
丁香假单胞菌	*P. syringae*		植物病原菌
根瘤菌			因其具有固氮活性而对于农业十分重要
荚膜红假单胞菌			光紫细菌
伤寒沙门氏菌	*S. typhi*		引起伤寒
鼠伤寒沙门氏菌	*S. typhimurium*	鼠伤寒血清型肠道沙门氏菌	常导致食物中毒,广泛用于遗传学研究
志贺氏菌			导致细菌性痢疾
金黄色葡萄球菌	*Staph. aureus*, *S. aureus*		医院内感染的主要原因
表皮葡萄球菌	*Staph. epidermidis*, *S. epidermidis*,		常见于皮肤组织,致病性弱于金黄色葡萄球菌
轻型链球菌	*Str. mitis*, *S. mitis*		口腔共生的机会致病菌
肺炎链球菌	*Str. pneumoniae*, *S. pneumoniae*,	肺炎球菌	革兰氏阳性球菌,导致肺炎和脑膜炎
化脓性链球菌	*Str. Pyogenes*, *S. Pyogenes*		会引起大范围的感染
链霉菌			革兰氏阳性细丝状土壤微生物,目前使用的抗生素的主要来源
霍乱弧菌	*V. cholerae*		因噬菌体编码毒素而导致霍乱
小肠结肠炎耶尔森氏菌	*Y. enterocolitica*		食源性传染的病原菌
鼠疫耶尔森氏菌	*Y. pestis*		导致鼠疫